

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 22:38:40 ; Search time 12 seconds  
(without alignments)  
1779.177 Million cell updates/sec

Title: US-10-098-602A-2

Perfect score: 2376

Sequence: 1 MTITYDELNLIRNGKIDTV.....INFEKQVTDWELNQGFNRY 454

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance: to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	480	20.2	439	1	GLNA-THEMA
2	466	19.6	443	1	GLNA-CLOSA
3	464	19.5	446	1	GLNA-LACIA
4	443	18.6	446	1	GLNA-METVO
5	442.5	18.6	443	1	GLNA-BACSU
6	424.5	17.9	446	1	GLNA-STAM
7	423	17.8	446	1	GLNA-STAM
8	412	17.3	445	1	GLNA-LACDE
9	411.5	17.3	443	1	GLNA-BACCE
10	400.5	16.9	446	1	GLNA-METMP
11	398.5	16.8	454	1	GLNA-METJA
12	395	16.6	442	1	GLNA-METTH
13	393.5	16.6	446	1	GLN2-MYCTU
14	391	16.5	443	1	GLNA-PYRKO
15	389.5	16.4	439	1	GLNA-PYRAB
16	389.5	16.4	439	1	GLNA-PYRFU
17	387.5	16.3	454	1	GLNA-HALVO
18	386.5	16.3	454	1	GLNA-HALNI
19	386	16.2	491	1	GLNA-ARCFU
20	381.5	16.1	439	1	GLNA-PYRHO
21	376	15.8	443	1	GLNA-PYRHO
22	353	14.9	435	1	GLN3-RHIME
23	341	14.4	472	1	YCJX-ECOLI
24	330.5	13.9	469	1	GLNA-AQURE
25	330	13.9	435	1	GLN3-RHILP
26	327	13.8	471	1	GLNA-SULSO
27	312	13.1	469	1	GLN1-STRRP
28	310.5	13.1	473	1	GLNA-ANASP
29	308.5	13.0	474	1	GLN1-FRAAL
30	303	12.8	478	1	GLN1-MYCTU
31	302	12.7	470	1	GLNA-FREDI
32	301	12.7	469	1	GLN1-STVRV
33	300.5	12.6	473	1	GLNA-SULAC

RESULT 1  
GLNA-THEMA  
ID AC P36205;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase) (GS).  
GN GLNA OR TM0943.  
OS Thermotoga maritima.  
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.  
OX NCBI\_TaxID=2336;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MSB8 / DSM 3109;  
RA MEDLINE=9228700; PubMed=1348781;  
RX Sanangelantoni A.M., Forlani G., Ambroselli F., Cammarano P., Tiboni O.;  
RA "The glnA gene of the extremely thermophilic eubacterium Thermotoga maritima: cloning, primary structure, and expression in Escherichia coli.";  
RT J. Gen. Microbiol. 138:383-393(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MSB8 / DSM 3109;  
RA MEDLINE=92287316; PubMed=10360571;  
RX Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A., McDonald L., Uitterback T.R., Malek J.A., Linher K.D., Garrett M.M., Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D., Heidelberg J., Sutton H.O., Venter J.C., Fraser C.M.;  
RA "Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima.";  
RL Nature 399:323-329(1999).  
CC -|- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate + L-glutamine.  
CC -|- SUBUNIT: Oligomer of 12 subunits arranged in the form of two hexagons (By similarity).  
CC -|- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -|- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; X60160; CAA42729.1; -  
EMBL; AE001758; AAD36024.1; -  
PIR; B72313; B72313.  
HSSP; P06201; ILGR.  
TIGR; TM0943; -  
InterPro; IPR001691; GLN\_synth.

#### ALIGNMENTS

34	299	12.6	469	1	GLNA-STRCO	P15106 streptomyc
35	293	12.3	473	1	GLNA-SYNP2	P28605 synechococ
36	287.5	12.1	469	1	GLN1-RHIME	Q59747 rhizobium m
37	287	12.1	472	1	GLNA-HAEIN	P43794 haemophilus
38	283	11.9	468	1	GLNA-METCA	P15124 methylococ
39	283	11.9	472	1	GLNA-PASMU	Q9C1P2 pasteurella
40	280	11.8	468	1	GLNA-AZOB	P10583 azospirillu
41	278.5	11.7	468	1	GLNA-VIBAL	P19204 vibrio algi
42	276	11.6	467	1	GLNA-AZOV	P22248 azotobacter
43	270	11.4	468	1	GLNA-AZOCA	P94126 azorhizobiu
44	269	11.3	481	1	GLNA-HELPI	P94845 helicobacte
45	267.5	11.3	469	1	GLNA-PROVU	P28786 proteus vul

*downscaled*

```

DR InterPro: IPR004809; GlnA.
DR InterPro: IPR001637; GlnA adenyltn.
DR Pfam: PF00120; gln-synt; 1.
DR Pfam: PF03951; gln-synt; 1.
DR ProDom: PD001057; Gln_synt_C; 1.
DR TIGRFAMs: TIGR00653; GlnA; 1.
DR PROSITE: PS00180; GlnA; 1.
DR PROSITE: PS00181; GlnA_ATP; 1.
KW Ligase; Complete proteome.
FT CONFLICT 204 280 A -> G (IN REF. 1).
FT CONFLICT 280 280 A -> R (IN REF. 1).
FT CONFLICT 336 336 S -> T (IN REF. 1).
SQ SEQUENCE 439 AA; 50035 MW; ABE3E674BD2F2359 CRC64;

Query Match 20.2%; Score 480; DB 1; Length 439;
Best Local Similarity 29.9%; Pred. No. 1.1e-30;
Matches 138; Conservative 73; Mismatches 209; Indels 42; Gaps 9;

QY 3 ITVDLNNLRNGKIDTVVACVDMOGRMGKLTGRHFLGLDQKKISISTFVYAVTIE- 61
DB 1 MTIETKRIEENVRIFRLOFTDINGTL-----KNLEITPDVLESWED 45

QY 62 GIAGGGEISS-VDTGYSDCHLCADLNSLHLLPWS-EGA-----VLAINPHNFTSEPL 114
DB 46 GIMFDGSSIEGFVRIEESDMYLPVLDTPAVLPTVDGAKSARVICDVTP-----DGKPF 101
QY 115 FCSFVILWQIIRLANLKLKGLFASLEPNLFNETHYKASQKHWNKLTQAQPHHWNMI 174
DB 102 EGDPRYLRRMKKAEQQLGYTPVAGPEMFFIIPINEKGEVPPEFLD-----HGGYFDL 155
QY 175 SASGGIETFMRSVRNKLKLEAGILMEATHPEFLPSQHELNFPVADPLTMADRIIAKHGVK 234
DB 156 LPLSKVEETRDITALEKMGITVEATHHEVAPSOHEVDFYDTFTRTADNAQTQKLVK 215
QY 235 EMAEQSGWATFMKLSSTALGNACHHMSLODAETKNAFYDQNDDEYGMSTLARNWIAG 294
DB 216 TMAIFGHYATFMKPKFYGVNGSMVHMSL---FRGDKNAFYDPPDPLGLSKELRYFVG 273
QY 295 LKXVPEATYFFASYINSYKRLQPLTFAPTKCWAIDNRTSAFRLCNSKSEGINVELRG 354
DB 274 ILKHAKALAAVTNPTINSYKRLVPGYEAAPVYISWSVNSALRIPKARGKATRLRYRGP 333
QY 355 GADLNPLAFSAIIAGISGIEKLELPPASGNVY-----DKELPEFPNSLQNAH 407
DB 334 DPCSNLYLAPAAIILAGLDGINKIEPPAPVEENIYHVSERREELNIESLPGLKEAVE 393
QY 408 LKESKMLNKTFGKILHYVNAANVEINEFSKQVTDWELNQ 449
DB 394 ELKKDDVIIDALGEHIFKFEVAAEKDWKEFSYVTINWELQR 435

RESULT 2
GLNA_CLOSA STANDARD; PRT; 443 AA.
AC P10656;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase).
GN GLNA.
OS Clostridium saccharobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OC NCBI_TaxID=169679;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P262;
RX MEDLINE=88086901; PubMed=2891680;
RA Janssen P.J., Jones W.A., Jones D.T., Woods D.R.;
RT "Molecular analysis and regulation of the glnA gene of the Gram-
RL positive anaerobe Clostridium acetobutylicum.";
RL J. Bacteriol. 170:400-408 (1988).
CC -I- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +

```

```

CC L-glutamine.
CC -I- ENZYME REGULATION: THE ACTIVITY OF THIS ENZYME IS CONTROLLED
CC BY ADENYLATION. THE FULLY ADENYLATED ENZYME COMPLEX IS INACTIVE.
CC -I- SUBUNIT: OLIGOMER OF 12 SUBUNITS ARRANGED IN THE FORM OF TWO
CC HEXAGONS.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
CC -I- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC -I- CAUTION: Was originally thought to originate from
CC C.acetobutylicum.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M18966; AA23241.1; -.
CC PIR: A28676; AJCLOA.
CC HSP: P06201; ILGR.
CC InterPro: IPR001691; Gln_synth.
CC InterPro: IPR004809; GlnA.
CC Pfam: PF001637; GlnA adenyltn.
CC Pfam: PF0120; gln-synt; 1.
CC Pfam: PF03951; gln-synt; 1.
CC ProDom: PD001057; Gln_synt_C; 1.
CC TIGRFAMs: TIGR00653; GlnA; 1.
CC PROSITE: PS00180; GlnA; 1.
CC PROSITE: PS00181; GlnA_ATP; 1.
CC PROSITE: PS00182; GlnA_ADENYLATION; FALSE_NEG.
KW Ligase
FT BINDING 392 392 AMP (UNDER CONDITIONS OF ABUNDANT
FT SEQUENCE 443 AA; 49682 MW; EDE5BF44495151FC CRC64;
SQ
Query Match 19.6%; Score 466; DB 1; Length 443;
Best Local Similarity 28.0%; Pred. No. 1.4e-29;
Matches 130; Conservative 84; Mismatches 211; Indels 40; Gaps 8;

QY 4 TYDELNNLRNGKIDTVVACVDMOGRMGKLTGRHFLGLDQKKISISTFVYAVTIEGI 63
DB 5 TKEDIINLVKGVKFIQLQFTDIFGLKVAITDKQL-----EKALDNECMFDGSSIDGF 60
QY 64 AGGGEIISVDTGYSDCHLCADLNSLHLLPW--SEGAVLAISNPHNFTSEPLFCSPRVI 121
DB 61 -----VRTESDMNLRLNLDSPVFPWRPQGGKVARLIDVYKPDGTFEGDPRHV 111
QY 122 LMQQIERLANLKLKGL-----FASLEFNLFNETHYKASQKHWNKLTQAQPHHWNISA 176
DB 112 L-----KANADAKELGYTMNVGPCEFFLE-----TDENGRATTNTQDKAGYFDLAP 160
QY 177 SSGIETFMRSVRNKLKLEAGILMEATHPEFLPSQHELNFPVADPLTMADRIIAKHGVREM 236
DB 161 TDLGENARRDMTALAEENGFEIASHHEVAEQNEIDPKYGDALTADNIMTFKLWVKSI 220
QY 237 AEQSGWATFMKLSSTALGNACHHMSLODAETKNAFYDQNDDEYGMSTLARNWIAGLL 296
DB 221 AQRHGLHASFPKPIFGINGSMHYVNSL--FKQGNAPVDNDKNGLSKVAYQIAGLL 278
QY 297 KVPPEATYFFASYINSYKRLQPLTFAPTKCWAIDNRTSAFRLCNSKSEGINVELRGGA 356
DB 279 KNIKMAAVTNPVNSYKRLVPGYEAAPVYLSCKNRTALLRVPARGAGTVELRCPPD 338
QY 357 DLNPLYAFSAIIAGISGIEKLELPPASGNVY-----NDKELPEFPNSLQNAHLL 409
DB 339 SSNPYLVLACLQAGLDGINKNLQPPAEVEANIFAMTEQERKENGIDNLPNNLYEAVNYM 398
QY 410 KESKMLNKTFGKILHYVNAANVEINEFSKQVTDWELNQENRY 454
DB 399 KENELAKKALGDHVYGVNVAAGKAAEWDYRTKVDHWELENYLUNKY 443

```

```
RESULT 3
ID   GLNA_LACLA          STANDARD;          PRT;          446 AA.
AC   Q9CDL9;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DE   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase).
GN   GLNA OR L12200.
OS   Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX   NCBI_TaxID=1360;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IL1403;
RX   MEDLINE=21235186; PubMed=1137471;
RA   Bolotin A., Wincker P., Mauer S., Jaillon O., Malarne K.,
RA   Weissenbach J., Ehrlich S.D., Sorokin A.;
RT   "The complete genome sequence of the lactic acid bacterium Lactococcus
RT   lactis ssp. lactis IL1403."
RL   Genome Res. 11:731-753(2001).
CC   -!- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +
CC   L-glutamine.
CC   -!- ENZYME REGULATION: DOES NOT SEEM TO BE REGULATED BY ADENYLATION.
CC   -!- SUBUNIT: Oligomer of 12 subunits arranged in the form of two
CC   hexagons (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
CC   EMBL; AE006449; AAK06298.1; -.
CC   PIR; H86899; H86899.
CC   HSSP; P06201; ILGR.
CC   InterPro; IPR001691; GLN synth.
CC   InterPro; IPR004809; GlnA.
CC   Pfam; PF00120; gln-synt; 1.
CC   ProDom; PD001057; Gln synt C; 1.
CC   TIGRFAMs; TIGR00653; GlnA; 1.
CC   PROSITE; PS00180; GlnA_1; 1.
CC   PROSITE; PS00181; GlnA ATP; 1.
CC   Ligase; Complete proteome.
CC   KW   Ligase; Complete proteome.
CC   SQ   SEQUENCE 446 AA; 49693 MW; COA953D5341AD5CF CRC64;

Query Match 19.5%; Score 464; DB 1; Length 446;
Best Local Similarity 23.0%; Pred. No. 2e-29;
Matches 137; Conservative 80; Mismatches 211; Indels 44; Gaps 11;

QY 1 MTITADIRRVKEDIKFELMTDILGTLKNEVPAATDQLKLENKMF-----GS 56
DB 1 MTITADIRRVKEDIKFELMTDILGTLKNEVPAATDQLKLENKMF-----GS 56
QY 59 TIEGIAGGYEISSVDGYSDCHLCADLNSLHLPWSE-----GAVLAISNPNFVTS 112
DB 59 TIEGIAGGYEISSVDGYSDCHLCADLNSLHLPWSE-----GAVLAISNPNFVTS 112
QY 57 SIEGF-----VRINESDMYLPDLDTWVFPWGDYGVKAVICDVVTP-----GE 103
DB 57 SIEGF-----VRINESDMYLPDLDTWVFPWGDYGVKAVICDVVTP-----GE 103
QY 113 PLFCSRVILMQOIERLANLKLGL-FASELEFNLP-NETYKSASQKHVKLTAOPHH 169
DB 113 PLFCSRVILMQOIERLANLKLGL-FASELEFNLP-NETYKSASQKHVKLTAOPHH 169
QY 104 PFAGDPGRVLKRNLSMEKLGFSNGLGPEPEFLKLNENDEPTLEVNKG----- 155
DB 104 PFAGDPGRVLKRNLSMEKLGFSNGLGPEPEFLKLNENDEPTLEVNKG----- 155
QY 170 QMNNIASSGIIETMRSVRNKELEAGILMEATHPEFLPSQHELFNVPADPLTMADRHIIA 229
DB 170 QMNNIASSGIIETMRSVRNKELEAGILMEATHPEFLPSQHELFNVPADPLTMADRHIIA 229
QY 156 GYFDLAPTDLAGNTRREIVNVLTDLGFEVEASHHEVAIGQHEIDFKVANALKACDNQIQF 215
DB 156 GYFDLAPTDLAGNTRREIVNVLTDLGFEVEASHHEVAIGQHEIDFKVANALKACDNQIQF 215
QY 230 KHGVREAEQSGMVATFMAKLSSTALGNACHIHMSLODAETEKNAFYDQNDYGNMSTLAR 289
DB 230 KHGVREAEQSGMVATFMAKLSSTALGNACHIHMSLODAETEKNAFYDQNDYGNMSTLAR 289

Db 216 KLVVKTITARKHGLHATFMAKPVHGINSGMHCHNSL-FTEDGANAFADPTGDMGLSDVAH 274
QY 290 NWIAGLLKYVPEATYFFASYNSYKRLQPLTFAPTKCCWAIDNRTSAFRLCNSKSEGINV 349
DB 275 SFIAGLLKHAYNTAINPTVNSYKRLVPGVEAPVYVWAGNRNRPILRVPASGLSTRV 334
QY 350 ELRIGGADLNPYLAFSAIIAGISIGIBEKLELPPASGNVY--NDKE-----LPEFPNSL 402
DB 335 ELRAVDPTANPYLALAVLLAAGLDGVBEKLEAPEAIESNIVYVMTTEERKAHGITDLPSTL 394
QY 403 QNATHLLKESKMLNKTEGKILHYHYNAANVEINEFSKQVTDWELNOGPNRY 454
DB 395 HNAVKALPEDIITVTEALGEHVLNVFNVEAKRIEWSAYAOQVFSOMEIDNLYLEY 446

RESULT 4
ID   GLNA_METVO          STANDARD;          PRT;          446 AA.
AC   P21154;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DE   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase) (GS).
GN   GLNA.
OS   Methanococcus voltae.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC   Methanococcaceae; Methanococcus.
OX   NCBI_TaxID=2188;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 1537 / PS;
RX   MEDLINE=90139872; PubMed=2575777;
RA   Possett O., Sibold L., Aubert J.-P.;
RT   "Nucleotide sequence and expression of the glutamine synthetase
RT   structural gene, glnA, of the archaeobacterium Methanococcus voltae."
RL   Res. Microbiol. 140:355-371(1989).
CC   -!- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +
CC   L-glutamine.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
CC   EMBL; X53509; CAA37585.1; -.
CC   PIR; A43995; A43995.
CC   HSSP; P06201; ILGR.
CC   InterPro; IPR001691; GLN synth.
CC   InterPro; IPR004809; GlnA.
CC   InterPro; IPR001637; GlnA adenyltn.
CC   Pfam; PF00120; gln-synt; 1.
CC   ProDom; PD001057; Gln synt N; 1.
CC   TIGRFAMs; TIGR00653; GlnA; 1.
CC   PROSITE; PS00180; GlnA_1; 1.
CC   PROSITE; PS00181; GlnA ATP; 1.
CC   Ligase.
CC   KW   Ligase.
CC   SQ   SEQUENCE 446 AA; 50199 MW; 95449E4DE8542690 CRC64;

Query Match 18.6%; Score 443; DB 1; Length 446;
Best Local Similarity 28.7%; Pred. No. 9.6e-28;
Matches 142; Conservative 73; Mismatches 182; Indels 98; Gaps 17;

QY 1 MTITYDELNLN--IRNGKID-----TVVLACVDMQGLMKRLTGRHFLGLDQKK 48
DB 6 MALEYIKNNVKLRFQFVDIHGEPKNIAYPVKLTADGGEELMGVLENGLFFDGG----- 60
```





RN SEQUENCE FROM N.A.  
RC STRAIN=Mu50 / ATCC 700699, and N315;  
RX MEDLINE=21311952; PubMed=11418146;  
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,  
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,  
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
RA Sekizizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,  
RA Kanihisa M., Yamasita A., Ohnima K., Furuya K., Yoshino C., Shiba T.,  
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;  
RT "Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*."  
RL Lancet 357:1225-1240 (2001).  
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +  
CC L-glutamine.  
CC -1- SUBUNIT: Oligomer of 12 subunits arranged in the form of two  
CC hexagons (by similarity).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AP003361; BAB57472.1; -  
CC EMBL; AP003133; BAB42404.1; -  
CC PIR; H89905; H89905.  
CC HSP; P06201; ILGR.  
CC InterPro; IPR001691; Gln\_synth.  
CC InterPro; IPR004809; GlnA.  
CC InterPro; IPR001637; GlnA\_adenyln.  
CC Pfam; PF00120; gln-synt; 1.  
CC Pfam; PF03951; gln-synt\_N; 1.  
CC ProDom; PD001057; Gln\_synt\_C; 1.  
CC TIGRFAMs; TIGR00653; GlnA; 1.  
CC PROSITE; PS00180; GlnA\_1; 1.  
CC PROSITE; PS00181; GlnA\_ATP; 1.  
CC Ligase; Complete proteome.  
CC  
CC SEQUENCE 446 AA; 50854 MW; FF9F25631167903D CRC64;  
  
Query Match 17.8%; Score 423; DB 1; Length 446;  
Best Local Similarity 28.5%; Pred. No. 3.8e-26;  
Matches 112; Conservative 82; Mismatches 169; Indels 30; Gaps 7;  
  
QY 78 SDCHLCADLNSLHLLPWS--EGAVLAISNPHNFTSEPLFCSPRVILMQOIERLANLKLK 135  
DB 68 SDMLHPDLDTWIFWFTACQGVKVARICDVYKDTGTPFGDPRANLKRVLKEMEDLQFT 127  
  
QY 136 GL-FASELEFNLFNETYKSAQHWKLNKLTAPQHHQ-----WNISASSGIETFMRSVR 188  
DB 128 DFNILGPEPEFFLKLDEK-----GEPTLELNDGGYFDLAPTDLGECRRDIV 175  
  
QY 189 NKLEBAGILMEATHPELPQCHELNFVPADPLTWADRHIIAKGVREMAEQSGWATFMA 248  
DB 176 LELEDMGFDIEAGHHEVAPQCHEIDFKYADAVTACDNIQTFKLWVKTIARKHNLHATFMP 235  
  
QY 249 KLSSTALGNACHTHMSLQDAETKNAFYQDNQDEYGVGMSSTLARNIAGLLKVVPEATYFFAS 308  
DB 236 KPLFGVNGSGMHNFSVL--FKGENAFDFTNTEWGLTETAYQTAGVLKQWAGFTAVCPN 293  
  
QY 309 YINSYKRLQPLTFAPTKCCWAIIDNRISAFRLCNKSEGINVELRIGADLNPLYAFSAII 368  
DB 294 LVNSYKRLVGYEAPCVIANGSKNRPLIRVPSRGLSTRIEVRSDPRAANPYMALAAAIL 353  
  
QY 369 AAGISGIEEKLELPPASGVNY--NDKE-----LPEFPNSLQNAATHLLKESKMLNTFGE 421  
DB 354 EAGLDGINKKLVKPEPVNQNIYEMNREEREAVGIQDLPTSLTYLTALKAMRENEVIKALGN 413

```

Db 158 YFDMASEDEGAKRREIVETLEKLGFRVEAAHVGDCQOQIDFRFDNALATADKLOTFK 217
Qy 231 HGVREMAQSQGVATFMAKLSSTALGNACHIHMSLQDAETEKNAFYQDNDEYGMSTLARN 290
Db 218 MVKTIARKYHLHASFMAKPVVEGLAGNGMHTNMSL--LKDGKNAFYDQGOYNLSTALT 275
Qy 291 WIAGLLKVPPEATYFFASYINSYKRLQPLTFAPTKCCWAIADNRTSAFRLCNKSEGINVE 350
Db 276 FLNGILEHARAITCVANPTVNSYKRLIPGFEAPYVISWASNRSPMVRIPNANEVGTGLE 335
Qy 351 LRIGGADINPYLASAITAAGISGIEE-KLELPPASGNVY--NDKELPE-----FNSL 402
Db 336 MRSTDPANPYLLLSACUKAGTGKEGLPMAPVTS--NLPEMTDDEKELGKLPSTL 394
Qy 403 QNATHLLKESKMLNKTFGEKLIHVVNAANVEINEFSKQVTDWELNQCQFN 452
Db 395 HNAIKAFKDEDEVKVSAFSEHIVDSFLELKETEWALYTSQVSEWEVKRYFN 444

RESULT 9
GLNA_BACCE STANDARD; PRT; 443 AA.
AC P19064;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase).
GN GLNA.
OS Bacillus cereus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1396;
RN (1)
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-18.
RX STRAIN=IFO 3131;
RA Nakano Y., Kato C., Tanaka E., Kimura K., Horikoshi K.;
RT "Nucleotide sequence of the glutamine synthetase gene (glnA) and its
upstream region from Bacillus cereus.";
RL J. Biochem. 106:209-215 (1989).
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +
L-glutamine.
CC -1- ENZYME REGULATION: NOT REGULATED BY POST-TRANSLATIONAL
MODIFICATION AND NOT SUBJECT TO FEEDBACK INHIBITION.
CC -1- SUBUNIT: OLIGOMER OF 12 SUBUNITS ARRANGED IN THE FORM OF TWO
HEXAGONS.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D00513; BAA00403.1; -.
CC PIR; J00075; AJBSQU.
CC HSP; P06201; ILGR.
CC InterPro; IPR001691; Gln_synth.
CC InterPro; IPR004809; GlnA.
CC InterPro; IPR001637; GlnA adenyltn.
CC Pfam; PF00120; gln-synt; 1.
CC Pfam; PF03951; gln-synt_N; 1.
CC ProDom; PD001057; Gln_synt_C; 1.
CC TIGRFAMs; TIGR00653; GlnA; 1.
CC PROSITE; PS00180; GlnA 1; 1.
CC PROSITE; PS00181; GlnA ATP; 1.
CC PROSITE; PS00182; GlnA_ADENYLATION; 1.
CC KW Ligase.
CC FT INIT_MET
SQ SEQUENCE 443 AA; 50064 MW; 14A9EAB948BE0657 CRC64;

```

```

Query Match 17.3%; Score 411.5; DB 1; Length 443;
Best Local Similarity 26.8%; Pred. No. 3.1e-25;
Matches 126; Conservative 82; Mismatches 212; Indels 51; Gaps 9;

Qy 4 TYDELANLIRNGKIDTVVLACVDMQGRMLGKRL-TGRHFLGLDOKKISISTFVAVTIEG 62
Db 4 TKEDIFFLAKEENVKYIRLQFTDILGLVKNVEIPVSQLTAKALDNKMMFDGS-----STEG 58
Qy 63 IAGGYEISSVDTSYSDCHLCADLNSLHLLPWS--EGAVLAISNPHNFVTSEPLFCSPRV 120
Db 59 F-----VRTEESDMYLPDLDTWVFPWTAEKVKVRLICDIYNADGTPPEGDPN 109
Qy 121 ILMOQIERLANLKLGL-FASELEFNLFNETYKSASQKHWKMLKTAQPHHOMNISASSG 179
Db 110 NLKRVKMEALGSDFNLGPEPEFFLF-----KVDEKGNPTLELNDNGG 154
Qy 180 I-----ETFMRSVRNKLKEAGILMEATHPEFLPSQHELNFVPADPLTMADRHIAK 230
Db 155 YFDLAPMDLGENCRDIVLELEENGFEIEASHHEVAPGQHEIDPKYANAIRSCDDIQTFK 214
Qy 231 HGVREMAQSQGVATFMAKLSSTALGNACHIHMSLQDAETEKNAFYQDNDEYGMSTLARN 290
Db 215 LVVKTIARKYHLHASFMAKPVVEGLAGNGMHTNMSL--FKNGENVFYDQGLQSLDDARH 272
Qy 291 WIAGLLKVPPEATYFFASYINSYKRLQPLTFAPTKCCWAIADNRTSAFRLCNKSEGINVE 350
Db 273 FIAGILKHAFAPTAVANPTVNSYKRLVPGVEAPCYVAWSAQNRSLVRIIPASRGISTRVE 332
Qy 351 LRIGGADINPYLASAITAAGISGIEE-KLELPPASGNVY-----NDKELPEFPNSLQ 403
Db 333 VRSVDPAANPYLVNATLAAAGLDGINKLTPPAADVNIYVMTKEEREAEAGIVDLPATLA 392
Qy 404 NATHLLKESKMLNKTFGEKLIHVVNAANVEINEFSKQVTDWELNQCQFNRY 454
Db 393 QALVTLOSNEVISNALGDHLEHFEAKFEPWFIDFTQVHQRKQDQMSLY 443

RESULT 10
GLNA_METMP STANDARD; PRT; 446 AA.
AC OS9648;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase) (GS).
GN GLNA.
OS Methanococcus maripaludis.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OX NCBI_TaxID=39152;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Lu;
RX MEDLINE=99084965; PubMed=9864338;
RA Cohen-Kupiec R., Marx C.J., Leigh J.A.;
RT "Function and regulation of glnA in the methanogenic archaeon
Methanococcus maripaludis.";
RL J. Bacteriol. 181:256-261 (1999).
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +
L-glutamine.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (potential).
CC -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF062391; AAD04845.1; -.

```



OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;  
OC Methanobacteriaceae; Methanothermobacter.  
OX NCBI\_TaxID=187420;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Delta H;  
RX MEDLINE=98037514; PubMed=9371463;  
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,  
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,  
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,  
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,  
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabakar S.,  
RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,  
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;  
RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
RT deltaH: functional analysis and comparative genomics.";  
RL J. Bacteriol. 179:7135-7155 (1997).  
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +  
CC L-glutamine.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
CC -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE000917; AAB86044.1; -  
DR PIR; F69076; F69076.  
DR HSSP; P06201; 1LGR.  
DR InterPro; IPR001691; Gln synth.  
DR InterPro; IPR004809; GlnA.  
DR InterPro; IPR001637; GlnA\_adenyltn.  
DR Pfam; PF00120; gln-synt; 1.  
DR Pfam; PF03951; gln-synt N; 1.  
DR PRODOM; PD001057; Gln\_synt C; 1.  
DR TIGRFAMs; TIGR00653; GlnA\_1.  
DR PROSITE; PS00180; GlnA\_1; 1.  
DR PROSITE; PS00181; GlnA\_ATP; 1.  
KW Ligase; Complete proteome.  
FT BINDING 372  
FT AMP (UNDER CONDITIONS OF ABUNDANT  
FT GLUTAMINE) (BY SIMILARITY).  
FT FT  
FT SEQUENCE 442 AA; 50248 MW; B907928AAD8960A6 CRC64;  
Query Match 16.6%; Score 395; DB 1; Length 442;  
Best Local Similarity 28.5%; Pred. No. 6.3e-24;  
Matches 115; Conservative 71; Mismatches 189; Indels 28; Gaps 8;  
QY 61 EGIAGGGYEISS-VDTGYSCHLCADLNSLHLLPW--SEGAVLAISNPHNFTVSEPLFCS 117  
DB 50 DGLLFDSGVGVFVINESDLVLPDPDTFTSTLPWRPEKGVCRFCIDIVPDKGPEGD 109  
QY 118 PRVILMOQIBRLANLKLGLFASLEFNLFNETYKVSASQKHWKLNKTAQPHH--QWNIS 175  
DB 110 PRVYLKALDKYAHGLGYEVNVPPEPFILDDQEDG-----NIIPDCGAYPDVE 159  
QY 176 ASSGIETFMRSVRNKLSEAGLMEATHPEFLPSQHELNFPVADPLTWADRHIIAKIGVRE 235  
DB 160 PVDQGTDFRKLVMNDLEALNFDVESHVAPGQHEIDFKALKATADAVITFKQA 219  
QY 236 MAEQSGNVATFMALSLSTALGNACHIHMSL-QDAETKNAFYQNDYGNSTLARWNIAG 294  
DB 220 IVDKIGVMVTFMPKPPFGENGSGHCHQSLSFKDGE---NVFYDPDTETQLSEALYFIGG 276  
QY 295 LLKVPVPATYFFASYINSYRKLQPLTFAPTKCWAIDNRTSAPFLCNSKSEGINVELRIG 354  
DB 277 LLKHAPALTAVCAPTVNSYKRLVPGYAPVYIAYGLKVRSLIRIPASRGKGTREVELRMP 336  
QY 355 GADLNPLYLAPSATIAAGISGIEKLELPPPPASGNVYNDKELPE-----FPNSLQONAT 406  
337 DPCNPFYLAFAAMLEAGMNGIQNKIDPGEPTIDVY-EKSMSELREMGITLPSLWEAY 395  
407 HLLKESKMLNKTGFKLILHYVNAANVEINFEKSKQVTDWELNQ 449  
396 HALEEDDVIKGALGGHVYKFEIKREWDYDVRVRFKYELEL 438  
RESULT 13  
ID GLN2 MYCTU STANDARD; PRT; 446 AA.  
AC Q10378; 1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Probable glutamine synthetase 2 (EC 6.3.1.2) (Glutamate--ammonia  
DE ligase 2).  
DE GLN2 OR RV2222C OR MT2280 OR MTCY427.03C OR MTCY190.33C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,  
RA Hornby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";  
RL Nature 393:537-544 (1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J.F., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains.";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +  
CC L-glutamine.  
CC -1- SUBUNIT: Oligomer of 12 subunits arranged in the form of two  
CC hexagons (By similarity).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z70692; CAA94645.1; -  
DR EMBL; AE0007073; AAK46565.1; -  
DR PIR; B07776; B07776.  
DR HSSP; P06201; 1LGR.  
DR TIGR; MT2280;  
DR Tuberculist; RV2222c;  
DR InterPro; IPR001691; Gln synth.  
DR InterPro; IPR004809; GlnA.  
DR InterPro; IPR001637; GlnA\_adenyltn.  
DR Pfam; PF00120; gln-synt; 1.  
DR Pfam; PF03951; gln-synt N; 1.



Search completed: December 17, 2003, 22:43:22  
Job time : 14 secs

```
RC STRAIN=GES / Orsay;
RX PubMed=12622808;
RA Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,
RA Poch O., Priour D., Querellou J., Ripp R., Thierry J.-C.,
RA Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.,
RT "An integrated analysis of the genome of the hyperthermophilic
RT archaeon Pyrococcus abyssi.";
RL Mol. Microbiol. 47:1495-1512(2003).
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +
CC L-glutamine.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ248288; CAB50513.1; --
DR PIR; C75009; C75009.
DR HSP; P06201; 11GR.
DR InterPro; IPR001691; Gln_synth.
DR InterPro; IPR004809; GlnA.
DR InterPro; IPR001637; GlnA_adenyln.
DR Pfam; PF00120; gln-synt; 1.
DR Pfam; PF03951; gln-synt_N; 1.
DR ProDom; PD001057; Gln_synt_C; 1.
DR TIGRFAMs; TIGR00653; GlnA; 1.
DR PROSITE; PS00180; GlnA_1; 1.
DR PROSITE; PS00181; GlnA_ATP; 1.
DR KW Ligase; Complete proteome.
FT BINDING 358 358 AMP (UNDER CONDITIONS OF ABUNDANT
FT GLUTAMINE) (BY SIMILARITY).
SQ SEQUENCE 439 AA; 49742 MW; 62CCFD370A98AF0 CRC64;

Query Match 16.4%; Score 389.5; DB 1; Length 439;
Best Local Similarity 28.0%; Pred. No. 1.7e-23;
Matches 128; Conservative 73; Mismatches 195; Indels 61; Gaps 15;

Qy 20 VLACVDMQGRMGKRLTGRHFLGLDQKKISISTFVAVVTIEGAGGYEISSVDTGYSD 79
Db 17 VQLIFVDINGMPKGEIPASRL-----QEAIE-----DGISFGSSVPGFQ-GIED 61

Qy 80 CHLC--ADLSLHLPRSEGAVLAISNPHNV--TSEPLFCSPRVILMQOIERLANLKLK 135
Db 62 SDLIFKADPDPTVEVPWDN-----VARYGYIYKGRPYGADPRGLVKRVIEKLAEMGIK 116

Qy 136 GLFASELEFNLFNETHYKSAQKHWK-NLKTAPHHQWNNISASSGIETFMESVRNKLSEA 194
Db 117 AYIGPEPEFYLFK---KNGS---WELEIPDVGGFDTLTDKADIK---REIAEYMPSP 167

Qy 195 GILMEATHPEFLPSQHELNFVPADPLTMADRHIIAKHGVRERMAEQSGMVAFTMAKLSSTA 254
Db 168 GLVPEVLHHEVGKQAQHEIDFRYDEALKTADNIIISFKYIVKAVAEVHGLYATFMKPIYGM 227

Qy 255 LGNACHTHMSLODAETEKNAFYQNDYGMSTLARNWITAGLLKVPETATYFFASYINSYK 314
Db 228 PGNGMHLHLSLW---KEGENIF---KGEGLSETALHFTGGLLKKAKALTATNPTVNSYK 282

Qy 315 RLQPLTFAPTCKCWAIDNRTSAFLCNKSGSINVELRIGGADINPVLAFSAIIAAGISG 374
Db 283 RLVPGEYAPVVISGYKNSALIRVPAPWNGARIERYCRDPDSANPYFAFMAILMAGLDG 342

Qy 375 IEKLELPPPPASGNVYNDKE-----LPEFPNSIQNATHLLKESKMLNKTFFGE----- 421
Db 343 IKHKVEPPAYVEENVYEMDEGKRKELGIDTLPGSLGEALDELEKDKVVRREALGEAYKNFI 402

Qy 422 -----KLILHYNNAANVEINEFSKVQTDWELNOGF 451
Db 403 EYKKEWESYLEYLEAKH--LPKDTKRVTEWELERYF 437
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 17, 2003, 22:51:35 ; Search time 2955 Seconds  
(without alignments)  
3734.089 Million cell updates/sec

Title: US-10-098-602A-2  
Perfect score: 2376  
Sequence: 1 MTITYDELNLRNGKIDTV.....INFSKQVTDWELNQGNRY 454

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO\_spool/US10098602/runat\_17122003\_151237\_28052/app\_query.fasta\_1.647  
-DB=EST -QPMI=fastap -SUFFIX=ret -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10098602@cgn 1.1 2810 @runat\_17122003\_151237\_28052 -NCPU=6 -ICPU=3  
-NO MWAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
EST:  
1: em\_estba:  
2: em\_esthum:  
3: em\_estin:  
4: em\_estmu:  
5: em\_estov:  
6: em\_estpl:  
7: em\_estro:  
8: em\_hc:  
9: gb\_est1:  
10: gb\_est2:  
11: gb\_hc:  
12: gb\_est3:  
13: gb\_est4:  
14: gb\_est5:  
15: em\_estfun:  
16: em\_estom:  
17: em\_gss\_hum:  
18: em\_gss\_inv:  
19: em\_gss\_pln:  
20: em\_gss\_vrt:  
21: em\_gss\_fun:  
22: em\_gss\_man:  
23: em\_gss\_mus:  
24: em\_gss\_pro:  
25: em\_gss\_rod:  
26: em\_gss\_png:  
27: em\_gss\_vrl:  
28: gb\_gss1:

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	777	32.7	501	10	AW870989 ra53c11.y
2	696	29.3	442	9	AW828772 ra73h03.y
3	506.5	21.3	621	14	CB374375 ru99e08.y
4	430.5	18.1	626	14	CB279008 ru39e11.y
C 5	395	16.6	850	29	BZ701245 PUBM151TD
C 6	389	16.4	825	29	CC443184 PUBH069TD
C 7	370	15.6	851	29	CC381873 PUBFA22TD
C 8	347	14.6	815	29	BZ705035 PUBM084TD
C 9	330.5	13.9	690	12	BZ347175 BJ347175
C 10	328	13.8	699	29	BZ684738 PUBFS60TD
C 11	321.5	13.5	1176	29	BZ569521 pac82-164
C 12	321	13.5	818	29	BZ687214 PUBD118TD
C 13	316.5	13.3	684	12	BZ358241 BJ358241
C 14	309.5	13.0	1138	10	BE366680 rockefell
C 15	295.5	12.4	584	9	AU039163 AU039163
C 16	293.5	12.4	883	29	CC439227 PUBM046TB
C 17	290.5	12.2	674	28	BH142677 TDGDR48TH
C 18	288	12.1	669	13	BH064742 Pgr_5_K19
C 19	285	12.0	682	29	BZ668122 PUBCR86TD
C 20	284	12.0	929	9	AA430929 M32 Where
C 21	279.5	11.8	705	14	CB970295 CAB10003
C 22	276.5	11.6	901	29	CC435159 PUBDF85TB
C 23	275.5	11.6	850	29	BZ704494 PUBM314TD
C 24	275.5	11.6	902	29	CC435165 PUBDF85TD
C 25	275.5	11.6	944	29	BZ679052 PUBG443TD
C 26	273	11.5	1124	29	BZ558114 pac81-60
C 27	271	11.4	676	10	BE431085 SUN011.B0
C 28	267.5	11.3	694	28	BH142971 TDGDU72TH
C 29	261	11.0	717	14	CB853860 UI-CF-DU1
C 30	260.5	11.0	727	9	AV928925 AV928925
C 31	251.5	10.6	818	12	BZ167896 BU167896
C 32	245.5	10.3	679	10	BE918734 FM1_1_H12
C 33	242.5	10.2	576	12	BM348177 MEST287-E
C 34	242.5	10.2	910	29	CC381870 PUBFA22TB
C 35	235.5	9.9	794	11	AV109282 Zea mays
C 36	235	9.9	619	9	AI896662 EST266105
C 37	233.5	9.8	564	12	BM140402 WHE0483.e
C 38	233.5	9.8	581	12	BZ268011 BJ268011
C 39	231	9.7	572	13	BU833379 T047C06.P
C 40	230.5	9.7	659	13	BU061228 FgrN_5_K1
C 41	230.5	9.7	698	13	BQ281957 WHE3027.C
C 42	230.5	9.7	698	13	BQ282096 WHE3028.G
C 43	230.5	9.7	743	14	CD452699 WHE1119.H
C 44	230.5	9.7	749	13	BU490570 604131271
C 45	230	9.7	487	13	BU276033 Cr_Emb_03

ALIGNMENTS

RESULT 1  
AW870989  
LOCUS  
DEFINITION  
ra53c11.y1 Bird-Rao Melloidogyne incognita J2 Melloidogyne incognita  
CDNA 5', similar to TR:088070.088070 PUTATIVE GLUTAMINE SYNTHETASE.  
/, mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
AW870989  
ra53c11.y1 Bird-Rao Melloidogyne incognita (southern root-knot nematode)  
EST.  
Melloidogyne incognita  
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;  
Tylenchoidea; Heteroderidae; Melloidogyninae; Melloidogyne.

```

REFERENCE
AUTHORS      1 (bases 1 to 501)
McCarte,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T.,
, Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.,
, Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagarisvill,R.,
, Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe
, M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.,
, Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
Wilson,R.
The Washington Univ. Nematode EST Project, 1999
Unpublished
Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was constructed by Uma Rao and David Bird
(david.bird@ncsu.edu) at North Carolina State University. DNA
Sequencing by: Washington University Genome Sequencing Center St.
Louis.
Seq primer: -40RP from Gibco
High quality sequence stop: 410.
Location/Qualifiers
1. .501
/organism="Meloiodogyne incognita"
/mol_type="mRNA"
/db_xref="taxon:6306"
/dev_stage="enriched for 2nd stage juveniles"
/lab_host="XLOL"
/clone_lib="Bird-Rao Meloiodogyne incognita J2"
/note="Vector: ZAP express - pBKCMV (Stratagene); Site 1:
EcoRI; Site 2: XhoI; Oligo (dT) primed library. cDNA was
constructed and cloned unidirectionally into the vector
within the 5' EcoRI and 3' XhoI sites. This library was
constructed by Dr. Uma Rao and Dr. David Bird at North
Carolina state University."
170 a 90 c 85 g 156 t
BASE COUNT
ORIGIN
Alignment Scores:
Pred. No.: 8.5e-79 Length: 501
Score: 777.00 Matches: 148
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 32.70% Indels: 0
DB: 10 Gaps: 0
US-10-098-602A-2 (1-454) x AW870989 (1-501)
QY 307 AlaSerTyrIleAsnSerTyrIlySerArgLeuGlnProLeuThrPheAlaProThrIlySerCys 326
Db 3 GCATCTTACATCACTCGTACAAAAGACTTCAACCGCTTACTTTTGCAGCAACAAAATGT 62
QY 327 CysTTPAlaIleAspAsnArgThrSerAlaPheArgLeuCysAsnSerIlySerGluGly 346
Db 63 TGTGGGCANTTGACACGACACAGCGCTTTCGACTTTGTAATTCAAATCCAGGGA 122
QY 347 IleAsnValGluLeuArgIleGlyAlaAspLeuAsnProTyrIleAlaPheSerAla 366
Db 123 ATTAATGTTGAGCTGGTATTGGTGGCGTGTATTTGAACCCCTATTAGCTTTTCCGCA 182
QY 367 IleIleAlaAlaGlyIleSerGlyIleGluGlyLeuLeuProProProAlaSer 386
Db 183 ATCATAGCTGCGAGGATTAAGCGGTATAGAAAGAAAGCTTGAACCTTCCCTCCTGCATCT 242
QY 387 GlyAsnValTyrAsnAspIlySerGluLeuProGluPheProAsnSerLeuGlnAsnAlaThr 406
Db 243 GGCATGTTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 302
QY 407 HisLeuLeuIlySerIlySerMetLeuAsnIlyThrPheGlyGlyLeuIleLeuHis 426
Db 303 CATCTTCTAAAGAAATCGAAATGCTGAATAAAACATTCGGGAGAGATTTGATTTCTACAT 362

```

---

```

QY 427 TyrValaenAlaAlaAsnValGluIleAsnGluPheSerLysGlnValThrAspTyrGlu 446
Db 363 TATGTAAACGCTGCTGAATGTTGAGATTAAATGAAATTTTCAAAACAAAGTTACTGACTGGAG 422
QY 447 LeuAsnGlnGlyPheAsnArgTyr 454
Db 423 CTTAATCAAGCATTTAATAGATAT 446
RESULT 2
AW828772 442 bp mRNA linear EST 10-MAY-2001
LOCUS ra73h03.Y1 Bird-Rao Meloiodogyne incognita J2 Meloiodogyne incognita
DEFINITION cDNA 5' similar to TR:088070 O88070 PUTATIVE GLUTAMINE SYNTHETASE.
; mRNA sequence.
ACCESSION AW828772
VERSION AW828772.1 GI:7922569
KEYWORDS EST.
SOURCE Meloiodogyne incognita (southern root-knot nematode)
ORGANISM Meloiodogyne incognita
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
Tylenchoidea; Heteroderidae; Meloiodogyninae; Meloiodogyne.
REFERENCE 1 (bases 1 to 442)
AUTHORS McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T.,
Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.,
Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagarisvill,R.,
Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe
, M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.,
Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
Wilson,R.
The Washington Univ. Nematode EST Project, 1999
Unpublished
Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was constructed by Uma Rao and David Bird
(david.bird@ncsu.edu) at North Carolina State University. DNA
Sequencing by: Washington University Genome Sequencing Center St.
Louis.
Seq primer: -40RP from Gibco
High quality sequence stop: 1.
Location/Qualifiers
1. .442
/organism="Meloiodogyne incognita"
/mol_type="mRNA"
/db_xref="taxon:6306"
/dev_stage="enriched for 2nd stage juveniles"
/lab_host="XLOL"
/clone_lib="Bird-Rao Meloiodogyne incognita J2"
/note="Vector: ZAP express - pBKCMV (Stratagene); Site 1:
EcoRI; Site 2: XhoI; Oligo (dT) primed library. cDNA was
constructed and cloned unidirectionally into the vector
within the 5' EcoRI and 3' XhoI sites. This library was
constructed by Dr. Uma Rao and Dr. David Bird at North
Carolina state University."
138 a 80 c 90 g 134 t
BASE COUNT
ORIGIN
Alignment Scores:
Pred. No.: 1.47e-69 Length: 442
Score: 696.00 Matches: 135
Percent Similarity: 93.20% Conservative: 2
Best Local Similarity: 91.84% Mismatches: 10
Query Match: 29.29% Indels: 0
DB: 9 Gaps: 0
US-10-098-602A-2 (1-454) x AW828772 (1-442)

```





Percent Similarity: 54.15% Conservative: 55  
 Best Local Similarity: 34.30% Mismatches: 115  
 Query Match: 16.62% Indels: 12  
 DB: 29 Gaps: 3

US-10-098-602A-2 (1-454) x BZ701245 (1-850)

QY 183 PheMetArgSerValArgAsnLysLeuGluAlaGlyIleLeuMetGluAlaThrHis 202  
 DB 849 TTCGCGCCACCGCATCGAGCGCTCGAGTCGATGCGATCTCGTCGAGTTCGCAC 790

QY 203 ProGluPheLeuProSerGlnHisGluLeuAsnValProAlaAspProLeuThrMet 222  
 DB 789 CACGAGGTGCCCCGGTTCAGCAGAGATCGATCTCCGTACGCGCGCTCGATG 730

QY 223 AlaAspArgHisIleAlaLysHisGlyValArgGluMetAlaGluGlnSerGlyMet 242  
 DB 729 GCCGACAACTGATCGCTTCGCTATCGTCAAGGAGTTCGATCGGTGAGGTG 670

QY 243 ValAlaThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHis 262  
 DB 669 TGGCGCAGTTCATGCCACAGCGCTCTCCGACCAACCGCGCTCGCGATGCACACGCAC 610

QY 263 MetSerLeuGlnAspAlaGluThrGluLysAsnAlaPheThrAspGlnAsnAspGluThr 282  
 DB 609 ATGAGCTGTTCGAGGGCGACAC-----AACGCTTTCACACACCGCGCGACCGCATG 556

QY 283 GlyMetSerThrLeuAlaArgAsnTrpIleAlaGlyLeuLysTyrValProGluAla 302  
 DB 555 CAGCTCTCGACACCGCAGCAGTTCATCGCGGATCTTCGACAGCGCGCGGATC 496

QY 303 ThrTyrPhePheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAla 322  
 DB 495 AGCGCGCTCACCAACAGTGGTGAACAGTTACAGCGCTTCATCCACGCGCGGAGGCA 436

QY 323 ProThrLysCysTyrTrpAlaLysAsnArgThrSerAlaPheArgLeu----- 339  
 DB 435 CCGACCGCGCGAGCTGGGTGGCGAAGCGCTTCGCGCTCATCTCGGCTCCGCTGTAC 376

QY 340 CysAsnSerLysSerGluGlyIleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsn 359  
 DB 375 ACGCCGAACAGGCGTCTCGCAGCATCGAGTTCGCGACGCCCGCTCGCTGTGTAAC 316

QY 360 ProTyrLeuAlaPheSerAlaIleAlaAlaGlyIleSerGlyIleGluGluLysLeu 379  
 DB 315 CCCTACCTGAGTTCGCGGTGCTCTCGCGCGGACTCAAGGCGCATCGCGAGATAC 256

QY 380 GluLeuProProAlaSerGlyAsnValTyr-----AsnAsp 392  
 DB 255 GAGCTGCGCGACGAGCGCGAGGACGCTCTGGGCGCTCACCTCGCGCGAGCTCGCACG 196

QY 393 LysGluLeuProGluPheProAsnSerLeuGlnAlaThrHisLeuLysGluSer 412  
 DB 195 ATGGGTACACCGAGTTCGCGCGCGGCTCGCGAGCGCTCAAGGCGATGGAGTCTCG 136

QY 413 LysMetLeuAsnLysThrPheGlyGluLysLeuIleLeuHisTyrValAsnAlaAlaAsn 432  
 DB 135 GAGTGTGTCGCGGCGACTCGCGGACGCTGTTCGACTACTTCTCTCCGCAACAGTGG 76

QY 433 ValGluLeuAsnGluPheSerLysGlnValThrAspTrpGluLeuAsnGln 449  
 DB 75 ACCGAGTGAACAGTACTACCGCAGCGCGGTTCACCGGTTCGAGTTCGCTCGC 25

## RESULT 6

CC443184/c  
 LOCUS  
 DEFINITION PUH069TD ZM 0.6-1.0 KB Zea mays genomic clone ZMMBTa453K17,  
 genomic survey sequence.  
 CC443184  
 ACCESSION  
 CC443184.1 GI:30946996  
 VERSION  
 GSS.  
 KEYWORDS  
 Zea mays  
 SOURCE  
 ORGANISM Zea mays

Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta,  
 Spermatophyta, Magnoliophyta, Liliopsida, Poales, Poaceae, PACCAD  
 clade, Panicoideae, Andropogoneae, Zea.  
 1 (bases 1 to 825)

REFERENCE  
 Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick  
 A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and Bennetzen, J.

AUTHORS  
 Maize Genomics Consortium

TITLE  
 Unpublished

JOURNAL  
 COMMENT

CONTACT: Cathy Whitelaw

712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TP

Class: sheared ends.

Location/Qualifiers

1..825

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="B73"

/db\_xref="taxon:4577"

/clone="ZMMBTa453K17"

/clone\_lib="ZM\_0.6-1.0 KB"

/notes="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high

/COT selected genomic DNA library"

BASE COUNT 121 a 258 c 298 g 148 t

ORIGIN

Alignment Scores:

Pred. No.: 8,83e-34 Length: 825

Score: 389.00 Matches: 93

Percent Similarity: 54.65% Conservative: 54

Best Local Similarity: 34.57% Mismatches: 110

Query Match: 16.37% Indels: 12

DB: 29 Gaps: 3

US-10-098-602A-2 (1-454) x CC443184 (1-825)

QY 191 LeuGluAlaGlyIleLeuMetGluAlaThrHisProGluPheLeuProSerGlnHis 210

DB 810 CTGAGTCGATGGGATCTCTCCGAGATCTTCGACACCGAGGGTCCCGGTGTCAGCAG 751

QY 211 GluLeuAsnPheValProAlaAspProLeuThrMetAlaAspArgHisIleAlaLys 230

DB 750 GAGATCGATCTCGCTACGCGCGAGCGCTGTCTGATGGCGCACACGCTCATGACCTTCGCG 691

QY 231 HisGlyValArgGluMetAlaGluGlnSerGlyMetValAlaThrPheMetAlaLysLeu 250

DB 690 TACATCGTCAAGGAGTTCGCGATCGGTGAGGTGTGTGGCGGACGTTTCATGTCACCAAGCCG 631

QY 251 SerSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeuGlnAspAlaGluThr 270

DB 630 CTCCTCGACACACCGCGCTCGGCGATGCACACATGAGCTGTTCGAGCGGCGACACC 571

QY 271 GluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMetSerThrLeuAlaAsn 290

DB 570 -----AACGCTTTCACACACCGCGAGCCCGATGCGAGCTCTCGGACACCGGCAACGAG 517

QY 291 TrpIleAlaGlyLeuLysTyrValProGluAlaThrTyrPhePheAlaSerTyrIle 310

DB 516 TTCATCGCGGATCTCTGACGACGCGCGGAGATCAGCGCGCTCACCACCAACGAGTGGGTG 457

QY 311 AsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThrLysCysTyrTrpAlaIle 330

DB 456 AACAGTTTACAGCGCTGATCCACGCGCGGCGAGCCCGACCGCGCGAGCTGGGGTGGC 397

QY 331 AspAsnArgThrSerAlaPheArgLeu-----CysAsnSerLysSerGluGlyIle 347

DB 396 GCGAACCGCTTCGCGCTCATCCGCTCCCGCTGTACACGCGCGAAGCGCTGCTCGCGCA 337

QY 348 AsnValGluLeuArgIleGlyGlyAlaAspLeuAsnProTyrLeuAlaPheSerAlaIle 367

```

Db      336 CGCATCGAGTCCGAGCCCGGAGTGGCTTGTAAACCCCTACCTGACGTTCCGGTGTG 277
Qy      368 lleAlaAlaGlyleSerGlyleGluGluLysLeuGluLeuProProAlaSerGly 387
Db      276 CTCGCGCGGAGTCAAGGCGATCCCGGAGAGTAGAGTCCCGAGCGAGCCGAGGAC 217
Qy      388 AsnValTyr-----AsnAspLysGluLeuProGluPheProAsn 400
Db      216 GAGCTGTGGCCCTCACCTCGCGGAGGTGTCACGAGGGGTACACGAGTGTCCCGCG 157
Qy      401 SerLeuGlnAsnAlaThrHisLeuLysGluSerLysMetLeuAsnLysThrPheGly 420
Db      156 AGCTCGCGGAGCGCTCAAGCGATGAGTCTCTCGGAGTGTCTCGCGGAGCACTCGGC 97
Qy      421 GluLysLeuLysHisTyrValAsnAlaAlaAsnValGluLysPheSerLys 440
Db      96 GAGCAGGTGTCGACTACTTCTCCGCAACAAAGTGAGCGAGTGAACGACTACCGCAGC 37
Qy      441 GlnValThrAspTyrGluLeuAsnGln 449
Db      36 CGGTCAACCCGTTGAGTCTCGTTCG 10

```

## RESULT 7

```

CC381873/3
LOCUS      CC381873
DEFINITION CC381873 851 bp DNA linear GSS 19-MAY-2003
            PUBM084TD ZM 0.6 1.0 KB Zea mays genomic clone ZMMBTA436D19,
            genomic survey sequence.

```

```

ACCESSION CC381873
VERSION   CC381873.1 GI:30859448
KEYWORDS  GSS.
SOURCE    Zea mays

```

## ORGANISM

```

Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.

```

## REFERENCE

```

1 (bases 1 to 851)
AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
            A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.

```

## TITLE

## JOURNAL

## COMMENT

```

Unpublished
Other_GSSs: PUHFA22TB
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.

```

## FEATURES

## source

```

1..851
    /organism="Zea mays"
    /mol_type="genomic DNA"
    /strain="B73"
    /db_xref="taxon:4577"
    /clone="ZMMBTA436D19"
    /clone_lib="ZM 0.6 1.0 KB"
    /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
    cot selected genomic DNA library"
BASE COUNT      121 a 269 c 308 g 153 t
ORIGIN

```

## Alignment Scores:

```

Pred. No.:      1,43e-31      Length:      851
Score:          370.00      Matches:      89
Percent Similarity: 54.30%      Conservative: 50
Best Local Similarity: 34.77%      Mismatches: 105
Query Match:      15.57%      Indels:      12
DB:              29          Gaps:      3

```

```

US-10-098-602A-2 (1-454) x CC381873 (1-851)

```

```

Qy      183 PheMetArgSerValArgAsnLysLeuGluGluAlaGlyleLeuMetGluAlaThrHis 202

```

```

Db      799 TTCCGCCCCACGCCATCGACGCGTGGAGTCGATGGGCATCTCCGTGCGATTCTCGCAC 740
Qy      203 ProGluPheLeuProSerGlnHisGluLeuAsnPheValProAlaAspProLeuThrMet 222
Db      739 CACGAGGTGCCCGGTCACGAGAGATCTCCCTACCGCCGCGCGCTGTCGATG 680
Qy      223 AlaAspArgHisIleAlaLysHisGlyValArgGluMetAlaGluGlnSerGlyMet 242
Db      679 GCCGACACGTCATGACCTTCGCTACATCGTCAAGGAGGTCCGATCGGTGAGGGTGTG 620
Qy      243 VallAlaThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHis 262
Db      619 TGGGCGAGCTTCATGCTCCCAAGCCGCTCTCCGACCAACGCGCGTCCGCGATGCACACGCAC 560
Qy      263 MetSerLeuGlnAspAlaGluThrLysAsnAlaPheTyrAspGlnAsnAspGluTyr 282
Db      559 ATGAGCTGTTCGAGGGGACAC-----AACGCTTCCACACCCCGAGACCCCGATG 506
Qy      283 GlyMetSerThrLeuAlaArgAsnTyrIleAlaGlyLeuLeuLysTyrValProGluAla 302
Db      505 CAGCTCTCGACACCGGCAAGCAGTTTCATCGCGGCATCTCGACGACGCGCGGAGATC 446
Qy      303 ThrTyrPheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAla 322
Db      445 AGCGCGCTCACCAACAGTGGGTGAACAGTTTACAAGCGCTGATCCACGCGCGCGGACGCA 386
Qy      323 ProThrLysCysTyrTyrAlaIleAspAsnArgThrSerAlaPheArgLeu----- 339
Db      385 CGACCGCGCGAGCGTGGGTGGCGCAACCGCTCCGCGCTCATCCGCGTCCCGCTGTATC 326
Qy      340 CysAsnSerLysSerGluGlyIleAsnValGluLeuArgIleGlyAlaAspLeuAsn 359
Db      325 ACGCCGACACAGGCGTCTCGCGAGCATCGAGTCCGAGCCCGAGCTCGGTTGTAAC 266
Qy      360 ProTyrLeuAlaPheSerAlaIleAlaGlyIleSerGlyIleGluGluLysLeu 379
Db      265 CCTACCTGAGCTTCCGCGTGTCTCGCGCGGACTCAAGGGCATCGCGGAGAGTAC 206
Qy      380 GluLeuProProProAlaSerGlyAsnValTyr-----AsnAsp 392
Db      205 GAGCTGCCCGAGGCGAGGACGACGTCTGGGCCCTCACCTCGCGCGAGCGTGCACG 146
Qy      393 LysGluLeuProGluPheProAsnSerLeuGlnAsnAlaThrHisLeuLysGluSer 412
Db      145 ATGGGCTACCGAGCTCCCGGAGCCTCGCGAGGCGCTCAAGGCCATGGAGTCTCTCG 86
Qy      413 LysMetLeuAsnLysThrPheGlyLysLeuLeuLeuHisTyrVal 428
Db      85 GAGCTGGTTCGCGAGGCACTCGCGAGCACGTTGTCGACTACTTCTC 38

```

## RESULT 8

## BZ705035/c

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## BZ705035

## PUBM084TD

## ZM 0.6 1.0 KB

## Zea mays

## genomic survey sequence.

## BZ705035

## BZ705035.1

## GI:28425443

## GSS.

## Zea mays

## Zea mays

## Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

## Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

## clade; Panicoideae; Andropogoneae; Zea.

## 1 (bases 1 to 815)

## Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick

## A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.

## Maize Genomics Consortium

## Unpublished

## Contact: Cathy Whitelaw

## TIGR

## 9712 Medical Center Drive, Rockville, MD 20850, USA

## Tel: 301-838-5843

Db	147	TTCTCTCCGCAACAAGTGGACCGACTACCGAGCGCGGTCA	CCCCGTCGAG	88
Qy	447	LeuAsnGln	449	
		:::	:::	
Db	87	CTCGGTGCG	79	
RESULT 9				
BJ347175/c				
LOCUS	BJ347175	630 bp	mRNA	linear
DEFINITION	BJ347175 Dictyostelium discoideum cDNA library, AF Dictyostelium			
	discoideum cDNA clone dda26g12 3', mRNA sequence.			

```

Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.
1 (bases 1 to 690)
Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
Full length cDNA of Dictyostelium discoideum at the aggregation
stage
Unpublished
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tschini@genes.nig.ac.jp.
Location/Qualifiers
1..690
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/cloned="dda26g12"
/sex="mat A"
/dev_stage="Aggregation stage"
/clone.lib="Dictyostelium discoideum cDNA library, AP"
BASE COUNT      204 a   110 c   112 g   262 t     2 others
ORIGIN
Alignment Scores:
Pred. No.:          3,59e-27           Length:             690
Score:              330.50            Matches:             77
Percent Similarity: 56.57%            Conservative:        35
Best Local Similarity: 38.89%         Mismatches:         81
Query Match:       13.91%             Indels:              5
DB:                  12                Gaps:               4
US-10-098-602A-2 (1-454) x BJ347175 (1-690)
QY    256 GlyAsnAlaCysHisIleHisMetSerLeuGlnAspAlaGluThrGluLysAsnAlaPhe 275
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    666 GGCTGCTCAGGACACATGCATCAAAATTTCATTTGTTTTAAAACC CGTAAGAACCTATTC 607
QY    276 TyrAspGlnAnanAspGluTyTGlyMetSerThrLeuAlaArgAsnTrpIleAlaGlyLeu 295
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    606 CTCGATGAATCCGCCCAACCATCATATCTGTCGATCTTCAAATTCATTCGTGCTGTCGCA 547
QY    296 LeuLysTrpValProGluAlaThrTyTPhePheAlaSerTyriLeAsnSertyrLYSArG 315
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    546 TTATTATTGTTGCNTGAGTTTTTACCATTTCGCACCAACAATCAATAGTTATAAACGT 487
QY    316 LeuGlnProLeuThrPheAlaProThrLyScysTirPalaleAspAsnArGrThrSer 335
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    486 TTATPCGATGGGTATTGGCCCAACCAACACCAACATGGGGAATGGTAATAGAACTGTA 427
QY    336 AlaPheArgLeuCysAnenSerLysSerGluGlyIleAsnValGluLeuArgIleGlyGly 355
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    426 GCATTAAAGAATTATCAAGGT---GGTAAGCACCAAGATCGAANTTCAGAGTCACAGA 370
QY    356 AlaAspLeuAnenProTyrLeuAlaPheSerAlaIleAlaAlaGlyIleSerGlyIle 375

```

```

Db      369 TCAGATGTAATCCCTTACATTTCAATTCGCGCTTCATTCGCTGCTGCTCTCTATGGTGC 310
Qy      376 GluGluLysLeuGluLeu---ProProAlaSerGlyAsnValTyrAan-----Asp 392
Db      309 ATTAACAATTAGATTAAACCAAAACCAATCATAGGTAAATAGTTATGATCTCTATATAA 250
Qy      393 LysGluLeuProGlu---PheProAsnSerLeuGlnAsnAlaThrHisLeuLeuLysGlu 411
Db      249 AAAGACTAGTTGAAGATTACCAAGATCATTCAGTGAATCAACTGAATTAATCAATAA 190
Qy      412 SerLysMetLeuAsnLysThrPheGlyGluLysLeuLeuHisThrValAsnAlaAla 431
Db      189 TCAAGATTGCAAAAGAAATACCTTGGTGAAGAATTGGTGAATTCGTTGAAATCTCT 130
Qy      432 AsnValGluLeuAsnGluPheSerLysGlnValThrAspTyrGluLeuAsnGln 449
Db      129 CGTTGGGAATATAGACAATTCATCAAGTTTCATAAATGGGAATTAGAAAGA 76

```

RESULT 10  
BZ684738/c  
LOCUS  
DEFINITION  
BZ684738 PUBF60TD ZM 0.6 1.0 KB Zea mays genomic clone ZMMBTA045123,  
genomic survey sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
BZ684738.1 GI:28243411  
Zea mays  
Zea mays

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
1 (bases 1 to 699)  
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick  
, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J., and Bennetzen, J.  
Maize Genomics Consortium  
Unpublished  
Contact: Cathy Whitelaw

TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5943  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: TF  
Class: sheared ends.

FEATURES  
source  
1..699  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBTA045123"  
/clone\_lib="ZM 0.6 1.0 KB"  
/notes="Vector: pCR4-toPO; Site 1: EcoRI; 0.6-1.0 kb high  
Cof selected genomic DNA library"

BASE COUNT  
ORIGIN  
98 a 218 c 252 g 131 t

Alignment Scores:  
Pred. No.: 7,11e-27 Length: 699  
Score: 328.00 Matches: 79  
Percent Similarity: 54.31% Conservative: 47  
Best Local Similarity: 34.05% Mismatches: 94  
Query Match: 13.80% Indels: 12  
DB: 29 Gaps: 3

US-10-098-602a-2 (1-454) x BZ684738 (1-699)

Qy 217 AlaAspProLeuThrMetAlaAspArgHisIleAlaLysHisGlyValArgGluMet 236

Db 697 GCAGACGCGCTGCGAGTCCGCAACAGTCATACCTTCGCTACATCGTCAGAGGTC 638

Qy 237 AlaGluGlnSerGlyMetValAlaThrPheMetAlaLysLeuSerThrAlaLeuGly 256

```

Db      637 GCGATCGGTGAGGTGTGTGGCGGACGTTTCATGCCAAAGCCGCTCTCCGACCAACCGCGC 578
Qy      257 AsnAlaCysHisIleHisMetSerLeuGlnAspAlaGluThrGluLysAsnAlaPheTyr 276
Db      577 TCGGCGATGCACACGACATGAGCTGTTCGAGGGCGGACAC-----AACCCCTTCAC 524
Qy      277 AspGlnAsnAspGluTyrGlyMetSerThrLeuAlaArgAsnTyrPileAlaGlyLeuLeu 296
Db      533 AACCCCGAGCCGATCGAGCTTCGGACACCGGCAAGCAGTTCATCGCCGGCATCTCTG 464
Qy      297 LysTyrValProGluAlaThrTyrPhePheAlaSerTyrIleAsnSerTyrLysArgLeu 316
Db      463 CAGCAGCGCGCGAGATCAGCGCGCTCACAACCAAGTGGGTGAACAGTTTACAAGGCGCTG 404
Qy      317 GlnProLeuThrPheAlaProThrLysCysCysTyrAlaIleAspAsnArgThrSerAla 336
Db      403 ATCCACGGCGGAGGACCGGACCGCGGAGTGGGTGGCGGCGGACCGCTCCCGGCTC 344
Qy      337 PheArgLeu-----CysAsnSerLysSerGluGlyIleAsnValGluLeuArgIle 353
Db      343 ATCCGGCTCCCGCTGTACACCGGCAAGGGCTGTCTCGGACGCGATCGAGTCCGCGC 284
Qy      354 GlyGlyAlaAspLeuAsnProTyrThrLeuAlaPheSerAlaIleAlaAlaGlyLysSer 373
Db      283 CCGGACTCGGCTGTAAACCCCTACCTGAGTTTCGGGTGCTGCTCGCGCGGAGTCAAG 224
Qy      374 GlyIleGluGluLysLeuLeuProProAlaSerGlyAsnValTyr----- 390
Db      223 GGCATCGCGGAGAACTACGAGCTGCCGACGAGCGGAGGAGCGTCTGGGCGCTCACC 164
Qy      391 -----AsnAspLysGluLeuProGluPheProAsnSerLeuGlnAsnAlaThr 406
Db      163 TCGGCGGAGCTCGGACGATGGCTACACCGAGCTGCCCGGAGGCTCGCCGAGGCGCTC 104
Qy      407 HisLeuLeuLysGluSerLysMetLeuAsnLysThrPheGlyGluLysLeuLeuHis 426
Db      103 AAGGCGATGGAGTCTCGGAGTGTGCGCGGAGGACACTCGGCGGAGCAGCAGTGTTCGACTAC 44
Qy      427 TyrValAsnAlaAlaAsnValGluLeuLeuGluPhe 438
Db      43 TTCTCTCCGCAACAGTGGACCGAGTGGAAACGACTAC 8

```

RESULT 11  
BZ569521/c  
LOCUS  
DEFINITION  
BZ569521 pac82-164\_871.s2 pac82-164 Pseudomonas aeruginosa genomic clone  
pac82-164\_871, genomic survey sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
BZ569521.1 GI:27204274  
Pseudomonas aeruginosa  
Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
1 (bases 1 to 1176)  
Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,  
Burns, J.L., Kaul, R. and Olsen, M.V.  
Whole-Genome-Sequence variation among multiple isolates of  
Pseudomonas aeruginosa library  
J. Bacteriol., (2002) In press  
Contact: Chris K. Raymond  
Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244  
Email: craymond@u.washington.edu  
Class: shotgun.

FEATURES  
source  
1..1176  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"

```

/dbstrain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164_871"
/clone_lib="pacs2-164"
/notes="clinical isolate 2-164 whole genomic shotgun
library."
BASE COUNT 221 a 334 c 366 g 255 t
ORIGIN

Alignment Scores:
Pred. No.: 9.22e-26 Length: 1176
Score: 321.50 Matches: 95
Percent Similarity: 46.52% Conservativeness: 52
Best Local Similarity: 30.06% Mismatches: 135
Query Match: 13.53% Indels: 34
DB: 29 Gaps: 8

US-10-098-602A-2 (1-454) x BZ569521 (1-1176)
Qy 141 GluLeuGluPheAsnLeuPheAsnGluThrTyrLysSerAlaSerGlnLysHisTrpLys 160
Db 936 GAGGTGAAGTTCTCTCAATCGGACTCAAGGGCA----- 898
Qy 161 AsnLeuLysThrAlaGlnProHisGlnTrpMetAsnLysSerAlaSerGlylle 180
Db 897 ACCCTTAAACGGGGCGCTCGATATGCCGCTG-----GTGCGGGCGATTTCTTGGG 844
Qy 181 GluThrPheMetArgSerValArg-----AsnLysLeuGlu 192
Db 843 AAAGGTTGGCAAGTCGTGCAAGTTTATTTCATCGGAGGACCTGGACGAATTACGT 784
Qy 193 -GluAla-----GlylleLeuMetGluAla 201
Db 783 CGAATGCTTCCAGGACTTCATCGACGGCGCGGCCCGAGGCTTCGCGCGGAAGCAAT 724
Qy 201 rHisProGluPheLeuProSerGlnHisGluLeuAsnPheValProLa---AspProLe 220
Db 723 CGTCGCGGAATCCGACCGCGGAGTTTCGAGGTCAACCTCAACACCGTCAACGACGCGCT 664
Qy 220 uThrMetAlaAspArgHisLleAlaLysHisGlyValArgGluMetAlaGluInSe 240
Db 663 CAAGGCTGCGACACCGCGGTGCTGCTCAAGCGGCTGCTCAAGAACATCGCTACGACCA 604
Qy 240 rGlyMetValAlaThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHi 260
Db 603 CGAGATGGACACCACTTCATGGCCAGCCCTATCCGGCCAGCGCCGCGGGAAGCACTGCA 544
Qy 260 sIleHisMetSerLeuGlnAspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAs 280
Db 543 CGTGCATATCTCGTGTCTGCAC---AAGCATGGCAACAACATCTTCCACGAGGATCC 487
Qy 280 pGluTyrGlyMetSerThrLeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyrValPr 300
Db 486 CGAG-----CAGAACCGCGCATTCGCCCATTCGCGGTGCTGCTCGAGACCTGCCC 433
Qy 300 oGluAlaThrTyrPheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuTh 320
Db 432 GGCCTCATGGCTTCTCTGCCGAGCTCACTCTACCGCGCTTCGGTTCGAGTT 373
Qy 320 rPheAlaProThrLysCysTyrTrpAlaIleAspAsnArgThrSerAlaPheArgLeuCy 340
Db 372 CTACGTGCCGAACCGCGAGCTGGGCGCTGACCAACCGCACCGCTGCGCGCTGCGCGTCC 313
Qy 340 sAsnSerLysSerGluGlylleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsnPr 360
Db 312 CACCGGACCGCGGAGCGGTGACCCCTGGAAATCATCGGGTGGCGGCTGCCGCGCAACCC 253
Qy 360 oTyrLeuAlaPheSerAlaIleAlaAlaGlyIleSerGlylleGluGluLysLeuGlu 380
Db 252 CTACTCTACTGCGAGCGGTGCTGGCAGGGCTTCATCAGCGGCTTACCAACAGAGTCCA 193
Qy 380 uLeuProProAlaSerGlyAsnValTyrAsnAspLysLeuLeuProGluPheProAs 400

```

```

192 GCCGGCGCCCGCGATCGAAGCAACTCTACGACGATGAG---CCGAGCCTCGCCGAA 136
400 nSerLeuGlnAsnAlaThrHisLeuLeuLysGluSerLysMetLeuAsnLysThrPheG1 420
135 CAATTGCGGAGCCCTGCGGAGCTGGACGAAGCGAGATCATCGGAGTACATCGA 76
420 yGluLysLeuLysHisTyrValAsnAlaAlaAsnValGluLle 435
75 CCCGAAGTACATCGACATCTCTGTCGCTCGAAGAAAGCGAGCTG 30

RESULT 12
BZ687214 818 bp DNA linear GSS 05-FEB-2003
LOCUS PUB118TD ZM 0.6-1.0 KB Zea mays genomic clone ZMMB7a029D11,
DEFINITION genomic survey sequence.
ACCESSION BZ687214
VERSION BZ687214.1 GI:28248163
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 818)
AUTHORS (Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
TITLE Maize Genomics Consortium
JOURNAL Unpublished
COMMENT Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
source 1..818
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMB7a029D11"
/clone_lib="ZM 0.6-1.0 KB"
/notes="Vector: pCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
COT selected genomic DNA library"
BASE COUNT 164 a 287 c 232 g 135 t
ORIGIN

Alignment Scores:
Pred. No.: 5.86e-26 Length: 818
Score: 321.00 Matches: 86
Percent Similarity: 49.25% Conservativeness: 45
Best Local Similarity: 32.33% Mismatches: 121
Query Match: 13.51% Indels: 14
DB: 29 Gaps: 3

US-10-098-602A-2 (1-454) x BZ687214 (1-818)
Qy 113 ProLeuPheCysSerProArgValIleLeuMetGlnIleGluArgLeuAlaAsnLeu 132
Db 36 CCCATGAACCTGTGCGCGCAACAGTCTGAAGAAGTCTCTGAGTGTGTACACCGACG 95
133 LysLeuLysGlyLeuPheAlaSerGluLeuGluPheAsnLeuPheAsnGluThrTyrLys 152
Db 96 GGTGCGCGGATCGTGGCGCGGAAATGAGTTTACCTG-----ACC 140
153 SerAlaSerGlnLysHisTyrLysAsnLeuLysThr-----Ala 165
Db 141 CAGCGTGGCAAGACCGGACCTCGCGCTGAAACCCACGATAGCCCTCGGGCGCGCC 200
166 GlnProHisGlnTrpMetAsnIleSerAlaSerGlylleGluThrPheMetArg 185

```

```

Db 201 GAAACCGGTCGCGAGTCGTTTCCATTGATGGCGGCAACGAATTCGACCCACTCTTCGAA 260
Qy 186 SerValArgAsnLysLeuGluGluAlaGlyIleLeuMetGluAlaThrHisProGluPhe 205
Db 261 GAGCTCTAGACTGGTGGCAAGCCCGGCGCTCGACACGTTGATCCACGAAGAC 320
Qy 206 LeuProSerGlnHisGluLeuAsnPheValProAlaAspProLeuThrMetAlaAspArg 225
Db 321 GCCCGCGGCACATGGAATCACTTCGCCACGGGATGCCCTCGACCTCGCTGACCCAG 380
Qy 226 HisIleAlaLysHisGlyValArgGluMetAlaGluGlnSerGlyMetValAlaThr 245
Db 381 ATCACTGTGTTCAAGCGCACCTCGCGAGGCGCTAAAGCAACACGTCGCCGCCACG 440
Qy 246 PheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeu 265
Db 441 TTCAATGCCCAAGCCGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 500
Qy 266 GlnAspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMetSer 285
Db 501 GTCGACATCCGACCGGCAACACGAGTTCGTCGACGCTGACGGC-----AACAGAGT 554
Qy 286 ThrLeuAlaArgAsnTrpIleAlaGlyLeuLysTyrValProGluAlaThrTyrPhe 305
Db 555 CGCTGTTCTCTGACACACATCGCGGCTTGCGAATAATACATCCCAAGCTGCTGCCGATG 614
Qy 306 PheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThrLys 325
Db 615 TTTCGCGCCCAACGTAACCTGTTCCGCTGCTTCCTGCGGACACTTCGGCACCCTGCAAC 674
Qy 326 CysCysTrpAlaIleAspAsnArgTyrSerAlaPheArgLeuCysAsnSerLysSerGlu 345
Db 675 GTCGAGTGGGGGAGAGAACCGCACCGTTCGCTGCGGCTGCGGCTGCGGCTGCGGCTG 734
Qy 346 GlyIleAsnValGluLeuArgIleGlyAlaAspLeuAsnProTyrTrpLeuAlaPheSer 365
Db 735 GCATCGCGGTGGAGAACGTTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 794
Qy 366 AlaIleAlaAlaGly 371
Db 795 GCTAGTCTGCTGTGTGGC 812

```

```

RESULT 13
BJ358241/c
LOCUS
DEFINITION
Dicoelostelium dicoelostelium cDNA clone dda5622 3', mRNA sequence.
ACCESSION
BJ358241
VERSION
BJ358241.1 GI:19257836
KEYWORDS
Dicoelostelium dicoelostelium
SOURCE
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE
1 (bases 1 to 684)
AUTHORS
Urushihara,H., Tanaka,Y., Kohata,Y. and Shin-i.T.
TITLE
Full length cDNA of Dictyostelium dicoelostelium at the aggregation
stage

```

```

Unpublished
Contact: Tadao Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
Location/Qualifiers
1. .684
/organism="Dictyostelium dicoelostelium"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dda5622"
/sex="mat A"

```

## FEATURES

```

source
1. .684
/organism="Dictyostelium dicoelostelium"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dda5622"
/sex="mat A"

```

```

/dev stage="Aggregation stage"
/clone lib="Dictyostelium dicoelostelium cDNA library, AF"
BASE COUNT 202 a 103 c 108 g 256 t 9 others
ORIGIN
Alignment Scores:
Pred. No.: 1.45e-25 Length: 684
Score: 316.50 Matches: 76
Percent Similarity: 56.12% Conservative: 34
Best Local Similarity: 38.78% Mismatches: 80
Query Match: 13.32% Indels: 6
DB: 12 Gaps: 4
US-10-098-602A-2 (1-454) x BJ358241 (1-684)
Qy 256 GlyAsnAlaCysHisIleHisMetSerLeuGlnAspAlaGluThrGluLysAsnAlaPhe 275
Db 669 GGCTGCTCAGGACCATCATCAAAATTTCAATTTTAAACACCGTAAAGAACNTATTTC 610
Qy 276 TyrAspGlnAsnAspGluTyrGlyMetSerThrLeuAlaArgAsnTrpIleAlaGlyLeu 295
Db 609 CTCGATGAATCCGACCCCAATCATATGCTGCATTTTCAATCATTCGTTGNTGTCGCA 550
Qy 296 LeuLysTyrValProGluAlaThrTyrPhePheAlaSerTyrIleAsnSerTyrLysArg 315
Db 549 TTATTATTGTTGCTGAGTTTTTACCATTTTTCGCCAACCAATCNATAGTTATAACGT 490
Qy 316 LeuGlnProLeuThrPheAlaProThrLysCysTyrTrpAlaIleAspAsnArgThrSer 335
Db 489 TTAGTCGATGGGTATTGGCCACCAACCAACATCGGGAGTGGATAATAGACTGTA 430
Qy 336 AlaPheArgLeuCysAsnSerLysSerGluGlyIleAsnValGluLeuArgIleGlyGly 355
Db 429 GCATTAAGAATTTATCAAAAGGT---GGTAAAGCAACAGATCAGATTTCAGATCAGGA 373
Qy 356 AlaAspLeuAsnProTyrTrpLeuAlaPheSerAlaIleIleAlaAlaGlyIleSerGlyIle 375
Db 372 TCAGATGTAATCCCTTACATTTCAATCGCGCTTCATTCGCTGCTGCTCTATGCTGTC 313
Qy 376 GluGluLysLeuGluLeu---ProProAlaSerGlyAsnValTyrAsn-----App 392
Db 312 ATTAACAAATTAGAATTTAAACAAACCAATCATAGGTAAATAGTATGATCTCTATAAA 253
Qy 393 LysGluLeuProGlu---PheProAsnSerLeuGlnAsnAlaThrHisLeuLysGlu 411
Db 252 AAAGACTAGTTGAAAGATTACCAAGATCATTTAGTGTATGATCACTCACTTACTATCAAAA 193
Qy 412 SerLysMetLeuAsnLysThrPheGlyGluLysLeuIleLeuHisTyrValAsnAlaAla 431
Db 192 TCAGATGTAATGCAAGATACCTTGTGAAGAATTTGTGATCATTTCTGTTGAAGACTCGT 133
Qy 432 AsnValGlu-IleAsnGluPheSerLysGlnValThrAspTrpGlu 446
Db 132 CGTTGGGAATATAGCACAAATTCATCATCAAGTTCTATAAATGGGAA 87
RESULT 14
BJ358241
LOCUS
DEFINITION
Dicoelostelium dicoelostelium cDNA clone dda5622 3', mRNA sequence.
ACCESSION
BJ358241
VERSION
BJ358241.1 GI:19257836
KEYWORDS
Dicoelostelium dicoelostelium
SOURCE
Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE
1 (bases 1 to 1138)
AUTHORS
Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Senses,C.W.,
Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
Philippe,H.
TITLE
The analysis of 100 genes supports the grouping of three highly
divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba

```

[illegible]

2

T A T G T C T G A C A T T T T C A A A T T C A T T C G T T G C T G G T C A A T A A T A T G T T G C C C T G A T T T T A A 60

2

```
Qy 304 TyrPhePheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAlaPro 323
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
61 CCATTTTCGCCACCAACATCAATAGTTATTAACGTTTAGTCGATGGTATGGGCACCA 120
Qy 324 ThrLysCysCysTyrAlaIleAspAsnArgThrSerAlaPheArgLeuCysAsnSerLys 343
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
121 ACCACCAACATGGGAATGGATAATAGAATCTAGCATTAGAATTTATCAAAAGGT--- 177
Qy 344 SerGluGlyIleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsnProTyrLeuAla 363
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
178 GGTAAAGCAACAAGATCAGATTCAGAGTCACAGGATCAGATGTAATCCTTACATTCA 237
Qy 364 PheSerAlaIleIleAlaAlaGlyIleSerGlyIleGluGluLysLeuGluLeu---Pro 382
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
238 ATCCCGCTTCATTCGCTCGCTGGTCTCATGCTGTGCTTACAATTAGAATTAACAACAA 297
Qy 383 ProProAlaSerGlyAsnValTyrAsn-----AspLysGluLeuProGlu---PhePro 399
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
298 AAACCAATCATAGTATAGTTATGATCTCTATAAAAAAGGACTAGTTGAAAGATTACCA 357
Qy 400 AsnSerLeuGlnAsnAlaThrHisLeuLeuLysGluSerLysMetLeuAsnLysThrPhe 419
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
358 AGATCATTAGCTGAATCAACTGAATTAATCAAAAATCAAAAGATTGCAAAAAGAAATACCTT 417
Qy 420 GlyGluLysLeuIleLeuHisTyrValAsnAlaAlaAsnValGluIleAsnGluPheSer 439
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
418 GGTGAAGAAATTTGTGATCAITTCGTTGAAACTCGTCGTTGGGAATATAGACAATTCAT 477
Qy 440 LysGlnValThrAspTyrGluLeuAsnGln 449
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
478 CATCAAGTTTCATAAATGGGAATTAGAAAGA 507
```

Search completed: December 18, 2003, 01:17:46  
Job time : 2974 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 18, 2003, 00:54:51 ; Search time 457 Seconds  
(without alignments)

3310.329 Million cell updates/sec

Title: US-10-098-602A-2

Perfect score: 2376

Sequence: 1 MTITYDELNNLRNGKIDTV.....INEFSKQVTDWELNGFNRY 454

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2211978 seqs, 1666101734 residues

Total number of hits satisfying chosen parameters: 4423956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US10098602/runat\_17122003\_151240\_28154/app\_query.fasta\_1.647  
-DB=Published\_Applications\_NA -Qfmt=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pcr -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -HEAPSZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US10098602 @CGN 1 1 353 @runat\_17122003\_151240\_28154  
-NCPU=6 -ICPU=3 -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT\_DSPBLOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US09\_PUB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
18: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
-----					

1	773.5	32.6	1362	15	US-10-156-761-6703	Sequence 6703, Ap
2	773.5	32.6	9025608	15	US-10-156-761-1	Sequence 1, Appli
3	475.5	20.0	5500	15	US-10-062-458-1	Sequence 1, Appli
4	470.5	19.8	1338	10	US-09-738-626-2442	Sequence 2442, Ap
5	470.5	19.8	3309400	10	US-09-738-626-1	Sequence 1, Appli
6	449	18.9	4259	9	US-09-733-383-12	Sequence 12, Appl
7	449	18.9	4259	13	US-10-199-405-12	Sequence 12, Appl
8	432	18.2	1347	9	US-09-815-242-9280	Sequence 9280, Ap
9	427.5	18.0	1359	15	US-10-156-761-5931	Sequence 5931, Ap
10	424.5	17.9	1338	9	US-09-815-242-4678	Sequence 4678, Ap
11	424.5	17.9	1341	9	US-09-815-242-8240	Sequence 8240, Ap
12	424.5	17.9	1341	9	US-09-815-242-8677	Sequence 8677, Ap
13	402	16.9	1356	9	US-09-815-242-6417	Sequence 6417, Ap
14	402	16.9	4176	10	US-09-070-927A-188	Sequence 188, App
15	395.5	16.6	1008	10	US-09-974-300-261	Sequence 261, App
16	328	13.8	722	10	US-09-974-300-4790	Sequence 4790, Ap
17	322	13.6	1431	10	US-09-738-626-3485	Sequence 3485, Ap
18	322	13.6	3309400	10	US-09-738-626-1	Sequence 1, Appli
19	300.5	12.6	918	15	US-10-156-761-200	Sequence 200, App
20	291	12.2	1407	15	US-10-156-761-5982	Sequence 5982, Ap
21	291	12.2	9025608	15	US-10-156-761-1	Sequence 1, Appli
22	287	12.1	1419	9	US-09-815-242-7040	Sequence 7040, Ap
23	287	12.1	1830121	15	US-10-329-960-1	Sequence 1, Appli
24	281.5	11.8	1503	15	US-10-187-267A-56	Sequence 56, Appl
25	281.5	11.8	36321	15	US-10-187-267A-1	Sequence 1, Appli
26	277	11.7	82993	16	US-10-080-170-645	Sequence 645, App
27	265.5	11.2	1410	9	US-09-815-242-7990	Sequence 7990, Ap
28	265	11.2	1446	9	US-09-815-242-7410	Sequence 7410, Ap
29	247	10.4	1410	9	US-09-815-242-9691	Sequence 9691, Ap
30	241	10.1	1410	9	US-09-815-242-6321	Sequence 6321, Ap
31	241	10.1	1410	13	US-10-299-799-2	Sequence 2, Appli
32	201	8.5	372	10	US-09-974-300-4799	Sequence 4799, Ap
33	159	6.7	330	10	US-09-974-300-4800	Sequence 4800, Ap
34	126.5	5.3	607	13	US-10-027-632-275319	Sequence 275319,
35	126.5	5.3	607	14	US-10-027-632-275319	Sequence 275319,
36	114.5	4.8	1419	15	US-10-205-823-151	Sequence 151, App
37	111.5	4.7	48715	13	US-10-238-075-921	Sequence 921, App
38	111.5	4.7	50538	13	US-10-085-959-57	Sequence 57, Appl
39	107	4.5	26668	10	US-09-962-832-222	Sequence 222, App
40	105	4.4	2793	10	US-09-917-800A-1629	Sequence 1629, Ap
41	104.5	4.4	2727	10	US-09-880-107-3748	Sequence 3748, Ap
42	102.5	4.3	858	10	US-09-974-300-1348	Sequence 1348, Ap
43	101	4.3	3567	15	US-10-101-464A-47	Sequence 47, Appl
44	100	4.2	5715	14	US-10-005-691-1	Sequence 1, Appli
45	100	4.2	6057	15	US-10-331-061-79	Sequence 79, Appl

#### ALIGNMENTS

RESULT 1  
US-10-156-761-6703  
; Sequence 6703, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 6703  
; LENGTH: 1362  
; TYPE: DNA

```

; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1362)
US-156-761-6703

```

**Alignment Scores:**

Pred. No.:	1,078-89	Length:	1362
Score:	773.50	Matches:	175
Percent Similarity:	56.6%	Conservative:	81
Best Local Similarity:	38.7%	Mismatches:	193
Query Match:	32.5%	Indels:	13
DB:	15	Gaps:	8

US-10-098-602A-2 (1-454) x US-10-156-761-6703 (1-1362)

Qy		6	AspGluLeuAasnAnenLeulleleluealgaandnglylsyleaspThrValValLeuAlaCysVal	25
Db		31	GAGGAGCTGCACGCCCTTCGCGGGCGCGAGATCGACA CTGTCTCTCGGTTCCTCC	90
Qy		26	AspMetGlnClArgLeuMetGlyVsbaArgLeuThrGlyArgHisPheLeuGlyLeuAsp	45
Db		91	GATTATCAAGGGCGCTTCAGGGCAAGCGGTTCGCCGCGCGTCTTCCTCTC	141
Qy		46	GlnLysLysIleSerIleSerThr-----PheValTyAlaValThrlleGlu	61
Db		142	GACGAGGTGTTGGAGCACGCCACGCGAGGATGTAAC TACTGTCTGCGGTTCGACCACCGAC	201
Qy		62	GlyllealaGlyGlyTyrrGlutIleSerSerValAspThrGlyTyrrSerAspCysHis	81
Db		202	ATGAACACCGTCGACGGCTATGACATCTCTCTCGGACCGGGGTACGGCATCTCGCC	261
Qy		82	LeuCysAlaAspLeuAasnSerLeuHisLeuLeuProTrpSerGluGlyAlaValLeuAla	101
Db		262	ATCATCCCCGACCTGAGCAGCTGCGCAGGTGCGTGTAACGCGGGCACGGCGCTGCTG	321
Qy		102	IleSerAasnProHiasnPheValThrSerGluProLeuPheCysSerProArgValIle	121
Db		322	ATCGCCGCAC---CTGCGCTTGAACACACGGATGCGCGGTGTGTGCGCGCCCCTCCGACATC	378
Qy		122	LeuMetGlnGIntleuArgLeuAlaAenLeuLysLeuLysGlyLeuPheAlaSerGlu	141
Db		379	CTGCGTGGCGAGCTGCACCGGTCTGCGCCCTCTGCGCTACACGCCCGCAGTTCGGCACCGAG	438
Qy		142	LeuGluPheAasnLeuPheAasnGluThrTyrrIysSerAlaSerGlnLysHisIstPrlysAasn	161
Db		439	CTGGAGTTCTATGCTCTTCAAGGACACCTACGACGAGCGCTGGGACGCGGTATCAAAGGG	498
Qy		162	LeuLysThrAlaclnProHisHisGlnTrpMetAenIleSerAlaSerSerGlylleGlu	181
Db		499	CTCACACCGGCCAACAGTACAAACATCGACTACTCTGGTGTCTGGGACGGGACGGATCGAA	558
Qy		182	ThrPheMetArgSerValArgAasnLysLeuGluAlaGlylleLeuMetGluAlaThr	201
Db		559	CCCCCTCTGCGCGCATCGCAACAGATGACGGCGCGGGCCCTCACCGTCGATCGGCC	618
Qy		202	HisProGluPheLeuProSerGlnHisGluLeuAenPheValProAlaAspProLeuThr	221
Db		619	AAGGGCGAGTGCAACCCCGCGCAGCAGAGATCGTGTTCGGGTACGACGAGGCCCTGGTC	678
Qy		222	MetAlaAspArgHisIleIleAlaLysHisGlyValArgGluMetAlaGluInSerGly	241
Db		679	ACCTGGCAGCAGCAGCGCGTCTCAAGAACC GGCGCCAGAGAGATCGCGCCAGGAGGGC	738
Qy		242	MetValAlaThrPheMetAlaLysLeuSerThrAlaLeuGlyAasnAlaCysHisIle	261
Db		739	GTCTCGATCACCTTCTATGCGCCAAGTACAACACGAGCGCGAG---GGCAACTCTCTGCCATC	795
Qy		262	HisMetSerLeuGlnAspAlaGluThrGluLysAenAlaPheTyrrAspGlnAasnAspGlu	281
Db		796	CACCTTCGTCTGCGGACCGCGAC---GGCACCAACGCCATGTGCGCGCGACGACCGCGC	852
Qy		282	TyrGlyMetSerThrLeuAlaargAsnTrpilleAlaGlyLeuLeuLysTyValPrGlu	301

## RESULT 2

```

1  ; Sequence 1, Application US/10156761
2  ; Publication No. US20030119018A1
3  ;
4  ; GENERAL INFORMATION:
5  ;
6  ; APPLICANT: OMURA, SATOSHI
7  ; APPLICANT: IKEDA, HARUO
8  ; APPLICANT: ISHIKAWA, JUN
9  ; APPLICANT: HORIKAWA, HIROSHI
10 ; APPLICANT: SHIBA, TADAYOSHI
11 ; APPLICANT: SAKAKI, YOSHIYUKI
12 ; APPLICANT: HATTORI, MASAHIRA
13 ;
14 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
15 ;
16 ; FILE REFERENCE: 249-262
17 ; CURRENT APPLICATION NUMBER: US/10/156,761
18 ; CURRENT FILING DATE: 2002-05-29
19 ; PRIOR APPLICATION NUMBER: JP 2001-204089
20 ; PRIOR FILING DATE: 2001-05-30
21 ; PRIOR APPLICATION NUMBER: JP 2001-272697
22 ; PRIOR FILING DATE: 2001-08-02
23 ; NUMBER OF SEQ ID NOS: 15109
24 ; SEQ ID NO 1
25 ; LENGTH: 9025608
26 ;
27 ; TYPE: DNA
28 ; ORGANISM: Streptomyces avermitilis
29 ;
30 ; FEATURE:
31 ;
32 ; NAME/KEY: misc feature
33 ; LOCATION: (4187715)
34 ; OTHER INFORMATION: a, t, c, g, other or unknown
35 ;
36 ; US-10-156-761-1

```

Alignment Scores:		
Pred. No.:	4.17e-83	Length:
Score:	773.50	Matches:
Percent Similarity:	56.64%	Conservative:
Best Local Similarity:	38.72%	Mismatches:
Query Match:	32.55%	Indels:
DB:	15	Gaps:
		8
		9025608

US-10-098-602A-2 (1-454) x US-10-156-761-1 (1-9025608)

Qy 6 AspGluLeuAsnLeuIleArgAsnGlyLysIleAspThrValValLeuAlaCysVal 25  
Db 8041067 GAGGAGCTGCAGCCCTCGTCGGCGCGGAGATCGACATGCTCTCGGTCC 8041126

Qy 26 AspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHisPheLeuGlyLeuAsp 45  
Db 8041127 GATATGCAAGCGCGCTCCAGGCAAGCGGTTCGCGCGCGCTTCTCCTC----- 8041177

Qy 46 GlnLysLysIleSerIleSerThr-----PheValTyrAlaValThrIleGlu 61  
Db 8041178 GACGAGGTGCTGGAGCAGCGGAGGAGATGCAACTACCTGCTCGCTCGACACCGAC 8041237

Qy 62 GlyIleAlaGlyGlyGlyTyrGluIleSerValAspThrGlyTyrSerAspCysHis 81  
Db 8041238 ATGAACACCGTCGACGGCTATGACATGCTCTCGGACCGCGGTACGGGACCTCGCC 8041297

Qy 82 LeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTyrSerGluGlyAlaValLeuAla 101  
Db 8041298 ATGCATCCCGACCTGAGCAGCTGCGCCAGGTGCGTGGAAACGCGGCGACGCGTCTG 8041357

Qy 102 IleSerAsnProHisAsnPheValThrSerGluProLeuPheCysSerProArgValIle 121  
Db 8041358 ATCCCGGAC---CTGCGCTGGAACGACGATCGCGGTGGTGGTGGTGGTGGTGGTGG 8041414

Qy 122 LeuMetGlnGlnIleGluArgLeuAlaAsnLeuLysLeuLysGlyLeuPheAlaSerGlu 141  
Db 8041415 CTGGTCCGACGCTGACCGGCTGCGCCCTCGCGTACACCGCCAGGTGCGGACCGGAG 8041474

Qy 142 LeuGluPheAsnLeuPheAsnGluThrTyrLysSerAlaSerGlnLysHisIleTyrLysAsn 161  
Db 8041475 CTGGAGTTTCATCGTTTCAAGGACACCTACGAGGAGCTGCGAGCGCGGATACAGGG 8041534

Qy 162 LeuLysThrAlaGlnProHisGlnThrPheMetIleSerAlaSerGlyIleGlu 181  
Db 8041535 CTCACCGGCGCAACACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8041594

Qy 182 ThrPheMetArgSerValArgAsnLysLeuGluAlaGlyIleLeuMetGluAlaThr 201  
Db 8041595 CCCTGTCGCGCGCATCCGCAAGATGAGTGGCGCGCGGCTCAGCGTGGTGGTGGTGG 8041654

Qy 202 HisProGluPheLeuProSerGlnHisGluLeuAsnPheValProAlaAspProLeuThr 221  
Db 8041655 AAGGCGAGTCAACCCCGGCGGACGAGATCGTGTTCGCTAGCAGGAGCGCTGCTGTC 8041714

Qy 222 MetAlaAspArgHisIleAlaLysHisGlyValArgGluMetAlaGluGlnSerGly 241  
Db 8041715 ACCTGCGACGACGCGCTCTACAAGACCGCGCGCAAGGAGATCGCGCGCGGAGGCG 8041774

Qy 242 MetValAlaThrPheMetAlaLysLeuSerThrAlaLeuGlyAsnAlaCysHisIle 261  
Db 8041775 GTGTCGATCATTATGGCCCAAGTACAAACGAGCGCGAG---GGCACTCTCGCCACATC 8041831

Qy 262 HisMetSerLeuGlnAspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAspGlu 281  
Db 8041832 CACCTGTGCTGGCGGACCGGAC---GGCACCAACGCTATGCGCGCGGACGCGCGCGG 8041898

Qy 282 TyrGlyMetSerThrLeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyrValProGlu 301  
Db 8041899 ---GGCATGTGCGAGCTCATGCGGCACTTCTCGCGGCGGAGCTCGCGCGCTGCGGCG 8041945

Qy 302 AlaThrTyrPhePheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPhe 321  
Db 8041946 TTCCTACTTGTATCGCGCCCAACATCACTCGTACAGCGGTTCACGCGGCTCTCTTC 8042005

Qy 322 AlaProThrLysCysCysTyrAlaIleAspAsnArgThrSerAlaPheArgLeuCysAsn 341  
Db 8042006 GCGCGGACCGCGTGGCTGGGTGAGTACGACACCGGACCTCGCGGCTGCGGCTGGTGC--- 8042062

Qy 342 SerLysSerGluGlyIleAsnValGluLeuArgIleGlyIleAlaAspLeuAsnProTyr 361  
Db 8042063 GGCCACGCGCGCTCGATCGCTCGAGAACCGGCTCCCGCGGTGAGTCAACCGCGAC 8042122

Qy 362 LeuAlaPheSerAlaIleAlaAlaGlyIleSerGlyIleGluGlyLeuGluLeu 381  
Db 8042123 CTCGCCGTCCGCGACTCGTCCGCGCGGCTGTACGCGCATCGAGCACAACTGGAGCTG 8042182

Qy 382 ProProAlaSerGlyAsnValTyrAsnAspLysGluLeuProGluPheProAsnSer 401  
Db 8042183 CCCGAGCGCTCGCGCGGCAACGCGCTAC---GCCCGGAGTACGAGCAGCTCCCGAC 8042239

Qy 402 LeuGlnAsnAlaThrHisLeuLeuLysGluSerLysMetLeuAsnLysThrPheGlyGlu 421  
Db 8042240 CTGCGGAGCGCGCGCAACTCTGGGAAACAGCCCAATCGGCGGCGCTTGGCGGAC 8042299

Qy 422 LysLeuIleLeuHisTyrValAsnAlaAlaAsnValGluIleAsnGluPheSerLysGln 441  
Db 8042300 GACGTGTCGCCCACTACCGCAACATCGCGCGCTCGACTGAGGCGCTTCGACGCGG 8042359

Qy 442 ValThrAspTrpGluLeuAsnGlnGlyPheAsnArg 453  
Db 8042360 GTGACCGACTGGGAGCTGCGCGCTCTTCGACGCG 8042395

RESULT 3

US-10-062-458-1  
; Sequence 1, Application US/10062458  
; Publication No. US20030003550A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAMURA, JUN  
; APPLICANT: IZUI, HIROSHI  
; APPLICANT: MORIGUCHI, KAYO  
; APPLICANT: NAKAMATSU, TSUYOSHI  
; TITLE OF INVENTION: METHOD FOR PRODUCING L-GLUTAMINE BY FERMENTATION AND L-GLUTAMIN  
; TITLE OF INVENTION: PRODUCING BACTERIUM  
; FILE REFERENCE: 219181USO  
; CURRENT APPLICATION NUMBER: US/10/062.458  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: JP 2001-28163  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: JP 2001-162806  
; PRIOR FILING DATE: 2001-05-30  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 1  
; LENGTH: 5500  
; TYPE: DNA  
; ORGANISM: Brevibacterium lactofermentum  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (659)..(1996)  
; OTHER INFORMATION:  
; FEATURE: CDS  
; NAME/KEY: CDS  
; LOCATION: (2006)..(5200)  
; OTHER INFORMATION:  
US-10-062-458-1

Alignment Scores:  
Pred. No.: 6,94e-50 Length: 5500  
Score: 475.50 Matches: 135  
Percent Similarity: 47.08% Conservative: 83  
Best Local Similarity: 29.16% Mismatches: 174  
Query Match: 20.01% Indels: 71  
DB: 15 Gaps: 12

US-10-098-602A-2 (1-454) x US-10-062-458-1 (1-5500)

Qy 15 GlyLysIleAspThrValValLeuAlaCysValAspMetGlnGlyArgLeuMetGlyLys 34  
Db 743 GGACACTTGAAGTCAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 793

Qy 35 ArgLeuThrGlyArgHisPheLeuGlyLeuAsnGlnLysLysIleSerIleSerThrPhe 54  
Db 793





```
QY 111 SerGluProLeuPheCysSerProArgValIleLeuMetGlnGlnIleGluArgLeuAla 130
Db 2363852 GGACACCCATCTTTTCTGACCGCGCCCAAGTCTCGCGAGGAGGTCCCACTAGCTGA 2363793
QY 131 AsnLeuLysLeuLysGlyLeuPheAlaSerGluLeuGluPheAsnLeuPheAsnGluThr 150
Db 2363792 GATGAAGGCTTGACCTGCATGATCTCACCAGAGATTGAGTTCTATTGTGGT----- 2363742
QY 151 TyrLysSerAlaSerGlnLysHisTrpLysAsnLeuLysThr----- 164
Db 2363741 -----CAAAGCCTTCGGACCAACCGAGTGCACCTGTG 2363709
QY 165 AlaGlnProHisGlnTrpMetAsnIleSerAlaSerSerGlyIleGluThrPheMet 184
Db 2363708 CCCACTGACACCGCGGATATTTCGACCAACCATTCATGAGGCGCGAATTTCCT 2363649
QY 185 ArgSerValArgAsnLysLeuGluAlaGlyIleLeuMetGluAlaThrHisProGlu 204
Db 2363648 CGAAACCGCATGCTGCTGGAGGAACCTCGCATCCCTGTCGAGTTCTCCACCATGAA 2363589
QY 205 PheLeuProSerGlnHisGluLeuAsnPheValProAlaAspProLeuThrMetAlaAsp 224
Db 2363588 ACTGCACCTGGCGCAGCAAGAAATTCGATTACGCCATCGCGATCGCTCACCATGGCCGAC 2363529
QY 225 ArgHisIleIleAlaLysHisGlyValArgGluMetAlaGluGlnSerGlyMetValAla 244
Db 2363528 AACATCATGACCTTCGCTACATCATGAACAGGTGCGAAGGACCAAGCGTCCGGGCA 2363469
QY 245 ThrPheMetAlaLysLeuSerSerThrAlaLeuLysAsnAlaCysHisIleHisMetSer 264
Db 2363468 TCATTTATGCCAAGCCATTCTCCAAAGCATGCGAGGTCCGCCATGCACACGACATGTCC 2363409
QY 265 LeuGlnAspAlaGluThrGluLysAsnAlaPheThrAspGlnAsnAspGluThrGlyMet 284
Db 2363408 TTATTTGGGGCGCATACC-----AACGCGTTCACGATCCAGACCATCTTACATGCTG 2363355
QY 285 SerThrLeuAlaArgAsnTrpIleAlaGlyLeuLysLysTrpValProGluAlaThrTyr 304
Db 2363354 TCCAAACCGCAAAACAGTTCATCGCTGGAATCTTCATCATCAGCTCCAGAAATTCACCGT 2363295
QY 305 PhePheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThr 324
Db 2363294 GTGACCAACACGATGGGTCAATTCCTACAAACGATCGTGTACGAAACGAAAGTCCCACT 2363235
QY 325 LysCysCysTrpAlaIleAspAsnArgThrSerAlaPheArgLeuCysAsn----- 341
Db 2363234 CGCGCAACCTGGGGGTATCTAATCGTCTCGCCTGCTGCTGCTTCTTCTACCTACCGTTTG 2363175
QY 342 SerLysSerGluGlyIleAsnValGluLeuArgIleGlyAlaAspLeuAsnProTyr 361
Db 2363174 AATAAGGAGGAGTCCGCGCGGGTGGAGGTGCGTCTTCTCCTCATACCGCTTGTAAACCATAT 2363115
QY 362 LeuAlaPheSerAlaIleIleAlaGlyIleSerGlyIleGluGluLysLeuGluLeu 381
Db 2363114 TTGGCGTTTTCAGTGATGCTCGCGCGTGTGTTTGAAGGCAATTAAGAGGTTATGAGTCT 2363055
QY 382 ProProAlaSerGlyAsnVal----- 389
Db 2363054 GACGACCGAGCTGAGACGATATCTCCAACTTGAGCTTCGGGAAACGTGCGGCCATGGGC 2362995
QY 390 TyrAsnAspLysGluLeuProGluPheProAsnSerLeuGlnAsnAlaThrHisLeuLeu 409
Db 2362994 TACAACGAT-----CTGCCAAGCAGCGCTTCATCAGGCACCTGCGCCCAATG 2362950
QY 410 LysGluSerLysMetLeuAsnLysThrPheGlyGluLysLeuIleLeuHisTyrValAsn 429
Db 2362949 GAAAGTCAGAGCTTGTGTGTCATCCTCGGTGAGCAGCTTTTTCAGTGTTCCTTTCGCGC 2362890
QY 430 AlaAlaAsnValGluLeuAsnGluPheSerLysGlnValThrAspTrpGluLeuAsnGln 449
Db 2362889 AATAAGTGGCGTGAATGGCGTGACTACCAAGACGATCATCTCCGTGGGAGCTCCGAAC 2362830
QY 450 GlyPheAsn 452
```

```
Db 2362829 AATCTTGAT 2362821
RESULT 6
US-09-733-383-12
; Sequence 12, Application US/09733383
; Patent No. US20010014467A1
; GENERAL INFORMATION:
; APPLICANT: Narbad, Arjan
; APPLICANT: Rhodes, Michael J. C.
; APPLICANT: Gasson, Michael J.
; APPLICANT: Walton, Nicholas J.
; TITLE OF INVENTION: PRODUCTION OF VANILLIN
; FILE REFERENCE: 20747/100
; CURRENT APPLICATION NUMBER: US/09/733,383
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: 09/155,183
; PRIOR FILING DATE: 1997-03-24
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 12
; LENGTH: 4259
; TYPE: DNA
; ORGANISM: Pseudomonas fluorescens
US-09-733-383-12
Alignment Scores:
Pred. No.: 1,258-46 Length: 4259
Score: 449.00 Matches: 128
Percent Similarity: 46.05% Conservative: 82
Best Local Similarity: 28.07% Mismatches: 224
Query Match: 18.90% Indels: 22
DB: 9 Gaps: 6
US-10-098-602A-2 (1-454) x US-09-733-383-12 (1-4259)
QY 1 MetThrIleThrTyrAspGluLeuAsnAsnLeuIleArgAsnGlyLysIleAspThrVal 20
Db 417 ATGAGTAACAACCTCGACCAAGTCTACCGATTTGTTGAAAGACCAACAAGATCACAAGATC 476
QY 21 ValLeuAlaCysValAspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHis 40
Db 477 GAATGCATGATTGGCGACTTGTACCGGATCACCCCGCGCAAGATCTCGCAACCAACAAG 536
QY 41 PheLeuGlyLeuAspGlnLysLysIleSerIleSerThrPheValTyrAlaValThrIle 60
Db 537 TTCAITTCGCAAAAAGGATGCGCTCCCGAGAGTGTCTGTTGCAGACAGTACGCGGC 596
QY 61 GluGlyIleAlaGlyGlyGlyTyrGluIleSerSerValAspThrGlyTyrSerAspCys 80
Db 597 GACTATGTGGAAGACGACATCTAT--TACGAACCTGCTCGACCCGCGCGACATCGACATG 653
QY 81 HisLeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTrpSerGluGlyAlaValLeu 100
Db 654 ATCTCGCGCGCGCGACAGAACGCGGTGTCTCTGTCATGGCCATCGAGCGCGACCGCG 713
QY 101 AlaIleSerAsnProHisAsnPheValThrSerGluProLeuPheCysSerProArgVal 120
Db 714 CAGGTGATTTCAGACACCTACGACAGCGGCAACCGCATCGAGTGTGCGCACCAAC 773
QY 121 IleLeuMetGlnGlnIleGluArgLeuAlaAsnLeuLysLeuLysGlyLeuPheAlaSer 140
Db 774 GTCCCTCAAGAAAGTCTCAAACTCTATTCCGACAAAGGGTGGCAGCGCATCGGCGCG 833
QY 141 GluLeuGluPheAsnLeuPheAsnGluThrTyrLysSerAlaSerGlnLysHisTrpLys 160
Db 834 GAAATGGAGTTCTACCTG-----ACCAAGCGCAGTGCAGCACCCTCGGATTAC----- 878
QY 161 AsnLeuLysThrAlaGlnPro-----HisHisGln 170
Db 879 -----CCATTGCAACCGCGGTGGCCCTTCGGACGTCGCGAATTCGTCGCCAA 929
QY 171 TrpMetAsnIleSerAlaSerSerGlyIleGluThrPheMetArgSerValArgAsnLys 190
```

Db 930 TCGTTCTATCGAAGCGCCCAAGAAATTCACCGCTGTTTCGAGACGCTACGACTGG 989  
Qy 191 LeuGluAlaGlyIleLeuMetGluAlaThrHisProGluPheLeuProSerGlnHis 210  
Db 990 TCGCAACTGCAGGAGCTGATCTCGATAGCTGATCCACGAGGAGCGCGCGCATG 1049  
Qy 211 GluLeuAsnPheValProAlaAspProLeuThrMetAlaAspArgHisIleAlaLys 230  
Db 1050 GAAATCAACTTCCTCGTCACGCGCGCGCTGTCCTCGCGGACCATCTCTGTTCAAG 1109  
Qy 231 HisGlyValArgGluMetAlaGluSerGlyMetValAlaThrPheMetAlaLysLeu 250  
Db 1110 CGCACCATGCGCGAGCGCGCTCAAGCAACGTCGCGCGCATCTATCGCAAGCGG 1169  
Qy 251 SerSerThrAlaLeuGluAlaAsnAlaCysHisIleHisMetSerLeuGlnAspAlaGluThr 270  
Db 1170 ATGACCGCGAGCTGCGAGCGCATGCACTGACACGAGCATCATCATGATGAGACC 1229  
Qy 271 GluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMetSerThrLeuAlaArgAsn 290  
Db 1230 GCGAAGAACGCTTTC-----TCCAATGAAGCGGAGCATGAGCCAGTTGTTCTCAAC 1283  
Qy 291 TrpIleAlaGlyLeuLysTyrValProGluAlaThrThrPhePheAlaSerTyrIle 310  
Db 1284 CACATCGCGCGCTGCGAATAATCATCCCTGAACGTGCTGCGCTGTTGCGGCCCAACGTC 1343  
Qy 311 AsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThrLysCysTyrAlaIle 330  
Db 1344 AACTCGTCCGCGCTCTCCGCGACACTTCGCGCGCGGTGACGTCGAGTGGGGGAA 1403  
Qy 331 AsnArgThrSerAlaPheArgLeuCysAsnSerLysSerGluGlyIleAsnValGlu 350  
Db 1404 GAAACCGTACCGTGGCGCTGCGGTGCGGATGCGCGCTCAAAACCGTGGGTGAA 1463  
Qy 351 LeuArgIleGlyAlaAspLeuAsnProTyrLeuAlaPheSerAlaIleAlaAla 370  
Db 1464 AACCGCTGCGCGGTGCGGACCAACCGTACCTGCGGATGCGCGAGCTGCTGTGC 1523  
Qy 371 GlyIleSerGlyIleGluGluLysLeuGluLeuProProAlaSerGlyAsnValTyr 390  
Db 1524 GGCTACATCGCATGTCGAGAGTATCAACCAAGCGCGCTGTTGGGTGCTGTTGTTAC 1583  
Qy 391 AsnAspLysGluLeuProGluPheProAsnSerLeuGlnAsnAlaThrHisLeuLys 410  
Db 1584 GAGCGCGCAACCTG---CGTCTCGCTGACCATCGAAGACGCTCTGGAACGATGAA 1640  
Qy 411 GluSerLysMetLeuAsnLysThrPheGlyGluLysLeuIleLeuHisTyrValAsnAla 430  
Db 1641 AACAGCAAGACCATCGAATAATCTCGGTGTCACAACTTCATCATCTGCTGCTGCTC 1700  
Qy 431 AlaAsnValGluIleAsnGluPheSerLysGlnValThrAspTrpGlu 446  
Db 1701 AAGCGGCGGAGCATGAATACTTCAAGCGGTGATCATGCTCATGGAA 1748

## RESULT 7

US-10-199-405-12

Sequence 12, Application US/10199405

Publication No. US20030167511A1

GENERAL INFORMATION:

APPLICANT: Narbad, Arjan

APPLICANT: Rhodes, Michael J.C.

APPLICANT: Gasson, Michael John

APPLICANT: Walton, Nicholas John

TITLE OF INVENTION: PRODUCTION OF p-HYDROXYBENZOIC ACID

FILE REFERENCE: 20747/103

CURRENT APPLICATION NUMBER: US/10199,405

CURRENT FILING DATE: 2002-07-17

PRIOR APPLICATION NUMBER: GB96/06187

PRIOR FILING DATE: 1996-03-23

PRIOR APPLICATION NUMBER: PCT/GB97/00809

PRIOR FILING DATE: 1997-03-24

PRIOR APPLICATION NUMBER: 09/155,183

;; PRIOR FILING DATE: 1998-09-22  
;; PRIOR APPLICATION NUMBER: 09/733,383  
;; PRIOR FILING DATE: 2000-12-07  
;; NUMBER OF SEQ ID NOS: 14  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 12  
;; LENGTH: 4259  
;; TYPE: DNA  
;; ORGANISM: Pseudomonas fluorescens  
US-10-199-405-12

## Alignment Scores:

Pred. No.: 1,25e-46 Length: 4259  
Score: 449.00 Matches: 128  
Percent Similarity: 46.05% Conservative: 82  
Best Local Similarity: 28.07% Mismatches: 224  
Query Match: 18.90% Indels: 22  
DB: 13 Gaps: 6

US-10-098-602A-2 (1-454) x US-10-199-405-12 (1-4259)

Qy 1 MetThrIleThrTyrAspGluLeuAsnAsnLeuIleArgAsnGlyLysIleAspThrVal 20  
Db 417 ATGATTAACAACCTCGACCACTGATGTTGTTGAAGACCAAGATCAGAGATC 476  
Qy 21 ValLeuAlaCysValAspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHis 40  
Db 477 GAATGTCATGATTGGCGACTTGGCGGATCACCAGCGGCAAGATCTCGCAACCAACAAG 536  
Qy 41 PheLeuGlyLeuAspGlnLysLysIleSerIleSerThrPheValTyrAlaValThrIle 60  
Db 537 TTCATTCCGCAAAAGGCGATCGCGCTGCCGAGAGTGTGCTGTTGAGACAGATGACGGC 596  
Qy 61 GluGlyIleAlaGlyGlyTyrGluIleSerSerValAspThrGlyTyrSerAspCys 80  
Db 597 GACTATGTCGAAGACGACATCTAT---TAGAACTGCTCGACCGCGCGCATCGACATG 653  
Qy 81 HisLeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTyrPheSerGluGlyAlaValLeu 100  
Db 654 ATCTGCGCGCGCGCGACGAGCGGTGTTCTCTGTCGATGGGCCATCGAGCGCGCG 713  
Qy 101 AlaIleSerAsnProHisAsnPheValThrSerGluProLeuPheCysSerProArgVal 120  
Db 714 CAGGTGATTCACGACACCTACGACGAGCGGCAACCCGATCGAGCTGTGCGCACGCAAC 773  
Qy 121 IleLeuMetGlnGlnIleGluArgLeuAlaAsnLeuLysLeuLysGlyLeuPheAlaSer 140  
Db 774 GTCTTCAAGAAAGTCTCAAACTCTATTCGCAAGGGCTGGCAGCGCGATCGTGGCGCG 833  
Qy 141 GluLeuGluPheAsnLeuPheAsnGluThrTyrLysSerAlaSerGlnLysHisTrpLys 160  
Db 834 GAAATGAGTTCTACCTG-----ACCAAGCGCAGTGCAGCGCGGATTAC----- 878  
Qy 161 AsnLeuLysThrAlaGlnPro-----HisGln 170  
Db 879 -----CCATTGCAACCGCGGTGGCGCTTCGCGAGCTCGGAAATCGGTGCGCAA 929  
Qy 171 TrpMetAsnIleSerAlaSerSerGlyIleGluThrPheMetArgSerValArgAsnLys 190  
Db 930 TCGTTCTCTATCGAAGCGCGCAAGAAATTCGACCCGCTGTTTGAAGACGCTCTACGACTGG 989  
Qy 191 LeuGluAlaGlyIleLeuMetGluAlaThrHisProGluPheLeuProSerGlnHis 210  
Db 990 TCGCAACTGCAGGAGCTGGATCTCGATACGCTATCCACGAGACGCGCGCGCATG 1049  
Qy 211 GluLeuAsnPheValProAlaAspProLeuThrMetAlaAspArgHisIleAlaLys 230  
Db 1050 GAAATCAACTTCCTGTCACGCGCGCGCTGTCCTCGCGCGACCATGATCTGTTGTTCAAG 1109  
Qy 231 HisGlyValArgGluMetAlaGluGlnSerGlyMetValAlaThrPheMetAlaLysLeu 250  
Db 1110 CGCACCATGCGCGAGCGCGCTCAAGCAACGTCGCGCGCATGCTGTTGTTGCGCAAGCG 1169

Qy	251	SerSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeuGlnAspAlaGluThr	270
Db	1170	ATGACCGCGCGACCTGGCGAGCGGCATGCACCTGCACACAGAGCATCATCGGATATCGGACG	1229
Qy	271	GluIysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMetSerThrLeuAlaAsn	290
Db	1230	GGC AAGAACGTCTTC-----TCCAAATGAAGACGGGAGCATGAGCCAGTGTTCCTCAAC	1283
Qy	291	TrpIleAlaGlyLeuLeuLysTyrValProGluAlaThrTyrPhePheAlaSerTyrIle	310
Db	1284	CACATCGCGGCGCTGCAGAAATTCATCCCTGAACCTGCTGCCGCTGTTCGGCCCAACGTC	1343
Qy	311	AsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThrLysCysCysTrpAlaIle	330
Db	1344	AACTCGTTCGCCCGCTTCCTGCCGGACACTTCGGCGCGGTGAACGTCGAGTGGGGCGAA	1403
Qy	331	AspAsnArgThrSerAlaPheArgLeuCysAsnSerLysSerGluGlyIleAsnValGlu	350
Db	1404	GAAACCGTACCGTGGCGCTTCGCGGTGCCGATGCCGGCCCTCAAAACCGTCGGGTGGAA	1463
Qy	351	LeuArgIleGlyAlaAspLeuAsnProTyrLeuAlaPheSerAlaIleIleAlaAla	370
Db	1464	AACGCTCGCGGTGCGAGCCCAACCCGTACTTCGCGATTCGCGGAGCCCTGCTGTC	1523
Qy	371	GlyIleSerGlyIleGluGluLysLeuGluLeuProProAlaSerGlyAsnValTyr	390
Db	1524	GGCTACATCGCATGGTCGAAGGTATCAACCAAGCGCGCTGTGTGGTGGTGGTGTTC	1583
Qy	391	AsnAspLysGluLeuProGluPheProAsnSerLeuGlnAsnAlaThrHisLeuLeuLys	410
Db	1584	GAGCGGGCGAACCTTG---CGTCTGCCCTCGACCATCGAAGACGCTTCGGAACGCATCGAA	1640
Qy	411	GluSerLysMetLeuAsnLysThrPheGlyGluLysLeuIleLeuHisTyrValAsnAla	430
Db	1641	AACAGCAGACCATCGAGAAATACCTGGGTACAACTTCATCATCTACTGCTAGCTCGCGGTC	1700
Qy	431	AlaAsnValGluIleAsnGluPheSerLysGlnValThrAspTrpGlu	446
Db	1701	AAGCGGCGCGAGCATGAAACCTTCAAGCGCGTGTATCAGCTCATCGGAA	1748

8. J. L. S. F. A.

RES001-8  
US-09-815-242-9280  
Sequence 9280, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA-011A  
CURRENT APPLICATION NUMBER: US/09/815.242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9280
; LENGTH: 1347
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1347)
US-09-815-242-9280

Alignment Scores:
Pred. No.: 2,79e-45 Length: 1347
Score: 432.00 Matches: 128
Percent Similarity: 44.80% Conservative: 83
Best Local Similarity: 27.18% Mismatches: 220
Query Match: 18.18% Indels: 40
DB: Gaps: 10

US-10-098-602A-2 (1-454) x US-09-815-242-9280 (1-1347)
Qy 1 MetThrIleThrTyAspGIuLeuAsnAsnLeuIleAArgAsnGlyLysIleAspThrVal 20
Db 1 ATGCCATCACAGCTGCAGATATTCGTCGTAAGTCAAGGAAAAAATGTTACCTTTATT 60
Qy 21 ValLeuAlaCysValAspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHis 40
Db 61 CGTCTTATGTTCTCAGATATTTTGGGAACCATGAAAAACGTCGAAATTCCT----- 111
Qy 41 PheLeuGlyLeuAspGln-----LysLysIleSerIleSerThrPheValTyAla 57
Db 112 -----GCTACAGATGAACAGTAGATGAAGGTTCTGTGCAACAAAGTTATGTTGATGGA 165
Qy 58 ValThrIleGluGlyIleAlaGlyGlyTyArgLysSerValAspThrGlyTy 77
Db 166 TCTTCTATTGAAGTTT-----GTACGTATCAATGAG 198
Qy 78 SerAspCysHisLeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTrpSerGlu--- 96
Db 199 TCGGATATGACTTGTATCCCGGACTTTCGATACATGGACAGCTCTCCCTGGGAGAGAA 258
Qy 97 --GlyAlaValLeuAlaIleSerAsnProHisAsnPheValThrSerGluProLeuPhe 115
Db 259 AATCGAAGTGTTCAGGCTCTGATCTGTGATGCTATACACAGAAAGTGGAACCATTTGCG 318
Qy 116 CysSerProArgValIleLeuMetGlnGlnIleGluArgLeuAlaAsnLeuLysLeuLys 135
Db 319 GGTGACCTCTGTGTAATTTGAAACGAGCTCTTCGTCACATGGNAGAGTTGGATTCAA 378
Qy 136 GlyLeu---PheAlaSerGluLeuGluPheAsnLeuPhe-----AsnGluThrTyLys 152
Db 379 TCCTTCAACCTTGTGTCAGAGCCAGAAATCTTCTCTATTAAAGTTGGATGAAATGGGAC 438
Qy 153 SerAlaSerGlnLysHisTrpLysAsnLeuLysThrAlaGlnProHisHisGlnTrpMet 172
Db 439 CCAACACTTGAAGTGAATGACAAAGGT-----GGCTACTTT 474
Qy 173 AsnIleSerAlaSerSerGlyIleGluThrPheMetArgSerValaArgAsnLysLeuGlu 192
Db 475 GATTGGCACCCTACTGACCTTGGCGGACACACACGCTCGTAGATGTGAAATGCTTGACC 534
Qy 193 GluAlaGlyIleLeuMetGluAlaThrHisProGluPheLeuProSerGlnHisGluLeu 212
Db 535 AAAATGGGATTTGAAGTAGAAGCGAGTCACACGAGGTTCGGGTGGACAGCATGACATT 594
Qy 213 AsnPheValProAlaAspProLeuThrMetAlaAspArgHisIleIleAlaLysHisGly 232
Db 595 GACTTTTAAGTACGATGAAGTTCTCCGTGCTTGTGATAAGATTCAAATCTTTAAGCTTGT 654
Qy 233 ValArgGluMetAlaGluGlnSerGlyMetValAlaThrPheMetAlaLysLeuSerSer 252
Db 655 GTTAAACCACTGTCTCCCAACACCGACTTATACGCAACNTTTATGGCGAGCCAAATTT 714
Qy 253 ThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeuGlnAspAlaGluThrGluLys 272
```

[illegible]

## RESULT 9

US-10-156-761-5931	Db	562	GGCGCGCCCGGCACGAGGAGATCGACTGGCTACGCGCAGCGCTCTCGACGCGGAC	6221
Sequence 5931, Application US/10156761				
Publication No. US20030119018A1	Qy	225	ArgHisIleIleAlaLysHisGlyValArgGluMetAlaGluGlnSerGlyMetValAla	244
GENERAL INFORMATION:				
APPLICANT: OMURA, SATOSHI	Db	622	AACATCATGACGTTCCGGCTGGTCATGAAGCAGGTGGCGCTGGAGCAGGCGGTCCAGCGC	681
APPLICANT: IKEDA, HARUO	Qy	245	ThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSer	264
APPLICANT: ISHIKAWA, JUN				
APPLICANT: HORIKAWA, HIROSHI	Db	682	ACCTTCATGCCGAGCGCTTCTCGGAGTACCCGGGCGACGGCGCATGCACACCCACTCTCTCG	741
APPLICANT: SHIBA, TADAYOSHI	Qy	265	LeuGlnAspAlaGluThrGluLysAsnAlaPheTyraSpGlnAsnAspGluTyGlyMet	284
APPLICANT: SAKAKI, YOSHIYUKI				
APPLICANT: HATTORI, MASAHIRO	Db	742	CTC-----TTCGAGGGCGACCGCAACGCGTCTTACGAGTCCGGCTCGGAGTACCAAGTC	795
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES				
FILE REFERENCE: 249-262	Qy	285	SerThrLeuAlaAraGAsnTrpIleAlaGlyLeuLeuLysTyraValProGluAlaThrTyr	304
CURRENT APPLICATION NUMBER: US/10/156,761	Db	796	TCCAAGTCCGGCGCTCTTCATCGCGCGCTCTGTAAGCAGCGGCGGAGATCTCGGCC	855
CURRENT FILING DATE: 2002-05-29	Qy	305	PhePheAlaSerTyriLeAsnSerTyriLysArgLeuGlnProLeuThr-----	320
PRIOR APPLICATION NUMBER: JP 2001-204089	Db	856	GTCAACCAACGAGTGGTCAACTCTCTACAAGCGCATCTGGGCGGCTCGGAGCGCACGCA	915
PRIOR FILING DATE: 2001-05-30	Qy	321	-----PheAlaProThrLysCysCysTrpAlaIleAspAsnArgThrSerAla	336
PRIOR APPLICATION NUMBER: JP 2001-272697	Db	916	GGCGCGCGGCGGAGGCTCCGTGCTATCTGCTGGGCGCCACAACACCGCTCCGCACTG	975
PRIOR FILING DATE: 2001-08-02	Qy	337	PheArgLeu-----CysAsnSerLysSerGluGlyIleAsnValGluLeuArgIle	353
NUMBER OF SEQ ID NOS: 15109	Db	976	GTCCGCGTCCCGATGTACAAGCCCGGCAAGACGGCTCGGCCCGCTCGAGGTCCGCTCC	1035
SEQ ID NO 5931				
LENGTH: 1359				
TYPE: DNA				
ORGANISM: Streptomyces avermitilis				
FEATURE:				
NAME/KEY: CDS				
LOCATION: (1)..(1359)				
US-10-156-761-5931				

QY 354 GlyGlyAlaAspLeuAsnProTyrLeuAlaPheSerAlaIleAlaAlaGlyIleSer 373  
Db 1036 CTGACTCGGCGCGAACCGTACCTGGCGTACGCGCTCTCTCGCGGCGCTCAAG 1095  
QY 374 GlyIleGluGlyLeuGluLeuProProAlaSerGlyAsnValTyr-----Asn 391  
Db 1096 GGCATCGAGGCGCTACGAGCTCCCGCGCGCGCGCGAGCGACGTGTGGCGCTCTCG 1155  
QY 392 AppLyGlu-----LeuProGluPheProAsnSerLeuGlnAenAlaThr 406  
Db 1156 GACGCGGAGCGCGCGATGGCATCGAGCCCTCGCGCAGAACCTCGCGAGGCGCTG 1215  
QY 407 HisLeuLeuGlySerLeuMetLeuAenLysThrPheGlyGlyLeuLeuLeuHis 426  
Db 1216 GCACCTATGCGCGCGAGCGCTGGCGCGCGAGCCTCGCGAGCGACGTCTTCGACTTC 1275  
QY 427 TyrValAsnAlaAlaAsnValGluIleAsnGluPheSerLysGlnValThrAspTrpGlu 446  
Db 1276 TTCCTCGCAACAGAGCAGGAGTGGAGGAGTACCGCTCGAGGTGACGCGCTTCGAG 1335  
QY 447 LeuAsnGln 449  
Db 1336 CTGCGGAAG 1344

## RESULT 10

US-09-815-242-4678  
; Sequence 4678, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4678  
; LENGTH: 1338  
; TYPE: DNA  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-4678

Alignment Scores:  
Pred. No.: 2,61e-44 Length: 1338  
Score: 424.50 Matches: 123  
Percent Similarity: 45.84% Conservative: 92  
Best Local Similarity: 26.23% Mismatches: 211  
Query Match: 17.87% Indels: 43  
DB: 9 Gaps: 9

US-10-098-602A-2 (1-454) x US-09-815-242-4678 (1-1338)

QY 2 ThrIleThrTyrAspGluLeuAsnAsnLeuIleAArgAenGlyLysIleAspThrValVal 21  
Db 13 ACTTTCACCTAAAGACGACATTCGTAATTTGCAGAGAGGAAATGTAGATATTAAAG 72  
QY 22 LeuAlaCysValAspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHisPhe 41  
Db 73 TTACAACTCACTGATATTATTAGAACAAATTAATAATGTTGAAGTGCCTGTAAAGCAATTA 132  
QY 42 LeuGlyLeuAspGlnLysLysIleSerIleSerThrPheValTyAlaValThrIleGlu 61  
Db 133 -----GAAAAAGTACTTGTATACGAAATGATGTTTCACGGTTCCTTCTATCGAA 180  
QY 62 GlyIleAlaGlyGlyTyrGluIleSerSerValAspThrGlyTyrSerAspCysHis 81  
Db 181 GGTTC-----GTACGTATCGAAGATTCAGATATGTATC 213  
QY 82 LeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTyrPheSer-----GluGlyAlaVal 99  
Db 214 TTACATCCAGATTAGATACCTGGGTAATCTTCCCATGGACTGCTGGACAAGTAAAGTT 273  
QY 100 LeuAlaIleSerAsnProHisAsnPheValThrSerGluProLeuPheCysSerProArg 119  
Db 274 GCACGTTTAAATTTGTGATGTATATAAAACAGATGGTACACCATTTTCAAGGGATCCTCGT 333  
QY 120 ValIleLeuMetGlnGlnIleGluArgLeuAlaAsnLeuLysLeuLysGlyLeu---Phe 138  
Db 334 GCAAACTTAAACGCTGTATTAAAGAAATGGAAGATTAGGCTTCACAGACTTTAACCTTA 393  
QY 139 AlaSerGluLeuGluPheAsnLeuPheAsnGluThrTyrLysSerAlaSerGlnLysHis 158  
Db 394 GGGCCTGGAACAGAAATCTCTTCTTTAAGTTGGATGAAAA----- 435  
QY 159 TrpLysAsnLeuLysThrAlaGlnProHisHisGln-----TrpMet 172  
Db 436 -----GGGAACCAACTTTAGAACTTAATGATGATGATGATGATATTC 477  
QY 173 AsnIleSerAlaSerSerGlyIleGluThrPheMetArgSerValArgAsnLysLeuGlu 192  
Db 478 GACTTAGCACCCTACAGATTAGGTGAAAACCTGCTGCGTATATTGTTTAGAATTAGAG 537  
QY 193 GluAlaGlyIleLeuMetGluAlaThrHisProGluPheLeuProSerGlnHisGluLeu 212  
Db 538 GATATGGGCTTCGATATTGAACCTAGTACCATGAAGTTGCCCTGGTCAACATGAAT 597  
QY 213 AsnPheValProAlaAspProLeuThrMetAlaAspArgHisIleIleAlaLysHisGly 232  
Db 598 GACTTTAAATATGCAGATGCTGTTACAGCATGTGATAACATCAACATTTAAATGGTT 657  
QY 233 ValArgGluMetAlaGluGlnSerGlyMetValAlaThrPheMetAlaLysLeuSerSer 252  
Db 658 GTTAAACCAATCGCACGTAACATATTTACACGCAACATTTATGCTTAACCATATT 717  
QY 253 ThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeuGlnAspAlaGluThrGlyLys 272  
Db 718 GGTGTGAATGGTAGCGGTATGCACITTAACGTTTCATTA-----TTCAAGAGTAAAGAA 771  
QY 273 AsnAlaPheTyrAspGlnAsnAspGluTyrGlyMetSerThrLeuAlaArgAsnTrpIle 292  
Db 772 AATGATTTCTTCGATCCAAATACTGAAATGGGCTTAACGGAACATGCATATCAATTTACA 831  
QY 293 AlaGlyLeuLeuLysTyrValProGluAlaThrTyrPhePheAlaSerTyrIleAsnSer 312  
Db 832 GCAGGTGTACTTAAATAATGCACGCGGATTTACTGCTGTATGTAACCCGTTAGTAACTCA 891  
QY 313 TyrLysArgLeuGlnProLeuThrPheAlaProThrLysCysCysTrpAlaIleAspAsn 332  
Db 892 TATAAACGTTTAGTACCTGGTTATGAAGCACCACATGTTATATGTCATGGAGTGGTAAAC 951  
QY 333 ArgThrSerAlaPheArgLeuCysAsnSerLysSerGluGlyIleAsnValGluLeuArg 352  
Db 952 CGTTCACCAATTAATCCGTGTACCATCTTCAAGAGGATTATCAACTCGTATCGAAGTACGT 1011

```

Qy 353 IleGlyAlaAspLeuAsnProTyrLeuAlaPheSerAlaIleAlaAlaGlyIle 372
Db 1012 TCAGTAGATCCAGCTGCAACCCATACATGCGTGTAGCTCAATCTTAGAAGCTGGACTA 1071
Qy 373 SerGlyIleGluGlyLeuLeuLeuProProAlaSerGlyAsnValTyr----- 390
Db 1072 GATGTTATTAATAATAAATTAAGTTCAGAACCAAGTTAACCAAAATATTTACGAAATG 1131
Qy 391 AsnAspLysGlu-----LeuProGluPheProAsnSerLeuGlnAsnAla 405
Db 1132 AACCGTGAAGACGTGAAGCAGTAGGATTCAGACTTACCTTCAACACTTTATACGTCA 1191
Qy 406 ThrHisLeuLeuLysGluSerLysMetLeuAsnLysThrPheGlyGluLysLeuIleLeu 425
Db 1192 TTAAGAGCAATGCGTGAATATCAAGTTATTAATAAAGCTTTAGGAATCATATCATATAAT 1251
Qy 426 HistyValAsnAlaAlaAsnValGluIleAsnGluPheSerLysGlnValThrAspTrp 445
Db 1252 CAATTTATTAATCAAAATCAATGATGAGGATTAACAGAACTCAAGATATCTGAATGG 1311
Qy 446 GluLeuAsnGlnGlyPheAsnArgTyr 454
Db 1312 GAAAGAGATCAGTACATCAACCAATAT 1338

```

## RESULT 11

```

US-09-815-242-8240
; Sequence 8240, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8240
; LENGTH: 1341
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1341)
US-09-815-242-8240

```

```

Alignment Scores:
Pred No.: 2,62e-44 Length: 1341
Score: 424.50 Matches: 123
Percent Similarity: 45.84% Conservative: 92
Best Local Similarity: 26.23% Mismatches: 211

```

```

Query Match: 17.87% Indels: 43
DB: 9 Caps: 9
US-10-098-602A-2 (1-454) x US-09-815-242-8240 (1-1341)
Qy 2 ThrIleThrTyrAspGluLeuAsnAsnLeuIleArgAsnGlyIleAspThrValVal 21
Db 13 ACTTTCATAAGACGACATTCGTAATTTTCGACAGAGGAGAAATGTAAGATATTTAAGA 72
Qy 22 LeuAlaCysValAspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHisPhe 41
Db 73 TTCAATTCCTGATATTTTAGGAACAAATTAATAATGTTGAAGTGCTGTGAAGCAATTA 132
Qy 42 LeuGlyLeuAspGlnLysLysIleSerIleSerThrPheValTyrAlaValIleGlu 61
Db 133 -----GAAAAAGTACTTGATAACGAAATGATGTTTGACGGTCTCTCTATCGAA 180
Qy 62 GlyIleAlaGlyGlyTyrGluIleSerSerValAspThrGlyTyrSerAspCysHis 81
Db 181 GGTTC-----GTACGTATCGAAGAATCAGATATGTATG 213
Qy 82 LeuCysAlaAspLeuAsnSerLeuHisLeuProTrpSer-----GluGlyAlaVal 99
Db 214 TTATCCAGATTTAGATACCTTGGTNAATCTTCCATGGACTGCTGACAGAGGTAAGATT 273
Qy 100 LeuAlaIleSerAsnProHisAsnPheValThrSerGluProLeuPheCysSerProArg 119
Db 274 GCAGGTTTAATTTGTGATGTATATAAACAGATGTTACACCATTTGAAGGGGATCCTCGT 333
Qy 120 ValIleLeuMetGlnGlnIleGluArgLeuAlaAsnLeuLysLysGlyLeu---Phe 138
Db 334 GCAAACTTAAACGTGTATTAAAGAAATGGAAGATTTAGGCTTCACAGACTTTAACTA 393
Qy 139 AlaSerGluLeuGluPheAsnLeuPheAsnGluThrTyrLysSerAlaSerGlnLysHis 158
Db 394 GGGCCTGAACCAAGATTTCTTGTGTTAAAGTTGGATGAAAAA----- 435
Qy 159 TrpLysAsnLeuLysThrAlaGlnProHisGln-----TrpMet 172
Db 436 -----GGGGAACCAACTTTAGAACTTAATGATGATGGTATTTTC 477
Qy 173 AsnIleSerAlaSerSerGlyIleGluThrPheMetArgSerValArgAsnLysLeuGlu 192
Db 478 GACTTAGACCTACAGATTTAGGTGAAACTGCTGCTGATATTTTGTAGAAATAGAG 537
Qy 193 GluAlaGlyIleLeuMetGluAlaThrHisProGluPheLeuProSerGlnHisGluLeu 212
Db 538 GATATGGGCTTCGATATTTGAAGCTAGTCACCAATGAAGTTGCCCTGCTCAACATGAAAT 597
Qy 213 AsnPheValProAlaAspProLeuThrMetAlaAspArgHisIleIleAlaLysHisGly 232
Db 598 GACTTTAAATATCGAGATGCTGTACAGCATGTGATAACATCCAAACATTAATTTGTT 657
Qy 233 ValArgGluMetAlaGluGlnSerGlyMetValAlaThrPheMetAlaLysLeuSerSer 252
Db 658 GTTAAACAAATCGACGTAAACATAATTTACACGCAACATTTATGCTAAACCATTTATTT 717
Qy 253 ThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeuGlnAspAlaGluThrGluLys 272
Db 718 GGTGTGAATGGTAGCGGTATGCACCTTTAAAGTTTCAATTA-----TTCAAGAGGTAAGAA 771
Qy 273 AsnAlaPheTyrAspGlnAsnAspGluTyrGlyMetSerThrLeuAlaArgAsnTrpIle 292
Db 772 AATGCATTTCTTCATCCAAATACTGAATGGGCTTAACGGAAACTGCATCATCAATTTACA 831
Qy 293 AlaGlyLeuLeuLysTyrValProGluAlaThrTyrPhePheAlaSerTyrIleAsnSer 312
Db 832 GCAGGTGTACTTAAAAATGCACGCGGATTTTACTGCTGTATGTAACCGCTTAGTAAACTCA 891
Qy 313 TyrLysArgLeuGlnProLeuThrPheAlaProThrLysCysCysTrpAlaIleAspAsn 332
Db 892 TATAAACGTTTAGTACCTGTTATGAAGCACCACCATGTTATATTCATCGATGGATGTTAAAC 951

```



```
Db      892 TATAAAGCTTTAGTACCTGGTTATGAAGCACCACCTATTATATTGTCATGGAGTGGTAAAC 951
Qy      333 ArgThrSerAlaPheAsgLeuCyAsnSerGluGlyIleAsnValGluLeuArg 352
Db      952 CGTTCACCAATTAATCCGTTGACCACTTCAAGAGATATCTACTCGTATCGAAGTACGT 1011
Qy      353 IlegGlyAlaAspLeuAsnProTyrLeuAlaPheSerAlaIleAlaAlaGlyIle 372
Db      1012 TCAGTAGATCCAGCTGCAACCCATACATACGCTGGTTAGTCAATCTTAGAGCTGGACTA 1071
Qy      373 SerGlyIleGluGlyLeuLeuLeuProProAlaSerGlyAsnValTyr----- 390
Db      1072 GATCGTATTAAATAATAAATAAAGTTCCAGAACCACTTAAACCAAAATATTACGAATG 1131
Qy      391 AsnAspGlyGlu-----LeuProGluPheProAsnSerLeuGlnAsnAla 405
Db      1132 AACCGTAGAAGACGTGAAGCAGTAGGCAATCAAGACTTACCTTCAACACTTATACGTCA 1191
Qy      406 ThrHisLeuLeuGlySerLysMetLeuAsnLysThrPheGlyGlyLeuLysLeuIleLeu 425
Db      1192 TTAAGAACGATGCGTGAAGATGAAGTTATTAAAGAGCTTTAGGAATCATATCTATAAT 1251
Qy      426 HisTyrValAsnAlaAlaAsnValGluLeuLeuGluPheSerLysGlnValThrAspTyr 445
Db      1252 CAATTTATTATTCAAAATCAATTTGAATGGATTACTACAGAACTCAAGTCTCTGAATGG 1311
Qy      446 GluLeuAsnGlnGlyPheAsnArgTyr 454
Db      1312 GAAAGAGATCATGATACATGAACCAATAT 1338
```

## RESULT 13

```
US-09-815-242-6417
; Sequence 6417, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6417
; LENGTH: 1356
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1356)
US-09-815-242-6417
```

---

Alignment Scores: 2.26e-41 Length: 1356  
Pred. No.: 402.00 Matches: 114  
Percent Similarity: 44.81% Conservative: 63  
Best Local Similarity: 28.86% Mismatches: 184  
Query Match: 16.92% Indels: 34  
DB: 9 Gaps: 7

US-10-098-602A-2 (1-454) x US-09-815-242-6417 (1-1356)

```
Qy      78 SerAspCysHisLeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTyrSerGluGly 97
Db      217 AGTGACATGATTATTATACCTGATCTATCTACATGATGATTTTTCGTGCGGAAGGCAC 276
Qy      98 -----AlaValLeuAlaIleSerAsnProHisAsnPheValThrSer 111
Db      277 CATGGAAAAGTTGCGCGCTTAATTTTGGCATATTATACATCT-----GATGGT 324
Qy      112 GluProLeuPheCysSerProArgValIleLeuMetGlnGlnIleGluArgLeuAlaAsn 131
Db      325 ACCCTTTTCCGCGAGATCCCGC-----GGTAAT 354
Qy      132 LeuLysLeuLysGlyLeuPheAlaSerGluLeuGluPheAsnLeuPheAsn----- 148
Db      355 TTAAGACGTGCTTTAGCAGATATCAAGAATAATAGGTTTACTCTCTTTTAACTTAGGACCA 414
Qy      149 -----GluThrTyrLysSerAlaSerGlnLysHisTrpLysAsnLeuLysThrAlaGln 166
Db      415 GAACCGAGATTTTTCCTATTATAATTAGATGAGTGAATGGAATTTACCACCTGATTGAAT 474
Qy      167 ProHisGlnTrpMetAsnIleSerAlaSerSerGlyIleGluThrPheMetArgSer 186
Db      475 GATAAAGTGGCTATTATTGATTTTGCACCAACCGATTAGTGGAATAATTGTCGCGGTGAT 534
Qy      187 ValArgAsnLysLeuGluGluAlaGlyIleLeuMetGluAlaThrHisProGluPheLeu 206
Db      535 ATTGTTTATAGATTAGAAAGTTTAGGCTTTGAAGTGAAGCTTCGCATCATGAAGTAGCA 594
Qy      207 ProSerGlnHisGluLeuAsnPheValProAlaAspProLeuThrMetAlaAspArgHis 226
Db      595 CTTGGTCAACATGAAATTTGATTTAAATATGCGGATGTAGTTGATGCTGTGCAATATT 654
Qy      227 IleIleAlaLysHisGlyValArgGluMetAlaGluGlnSerGlyMetValAlaThrPhe 246
Db      655 CAAACGTTTAAGTTAGTTTGGAATAATTTGCGCAACACATGTTTGCATGCACACATTT 714
Qy      247 MetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeuGln 266
Db      715 ATGCCAAAACCTTTGTTGGAAATTAATGTTCTGTTATGTCACATGTCATATGTCAATTA 771
Qy      267 AspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMetSerThr 286
Db      772 ---TTTAATGAAGAAGGTAATGTTTTTATGACGAGTTCAGGAGAGATGGGCTTTAAGCAA 828
Qy      287 LeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyrValProGluAlaThrTyrPhePhe 306
Db      829 ACTGCTTATCATTTCTTAGTGGTTTATTAAACATGCCCGCTGCCTATATCTCGGTATGT 888
Qy      307 AlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThrLysCys 326
Db      889 AACCCACAGTGAACCTTATAAACGCTTGGTACCAGGTTTATGAAGCACCTGTTTATGTA 948
Qy      327 CysTrpAlaIleAspAsnArgThrSerAlaPheArgLeuCysAsnSerLysSerGluGly 346
Db      949 GCTTGGAGCGGAGAAATCGTTCCACCATCTCATTCGAGTTCTCTGTTTATTCG 1008
Qy      347 IleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsnProTyrLeuAlaPheSerAla 366
Db      1009 ACTCGTTTAGAATACGTTCTGTCGATCCCTTACGCAATCCGTTATTAAACATGGCGATC 1068
Qy      367 IleIleAlaAlaGlyIleSerGlyIleGluGlyLeuLeuLeuProProAlaSer 386
```

```

Db      1069 TTATTACAAGCTGTTGGATGATTCGTAATGAACCTTACGCCACCGCAGCGATTGAT 1128
Qy      387 GlyAsnValTyr-----AsnAspLysGlu-----LeuProGluPhePro 399
Db      1129 CGTAATATTATGATGATGAGGAGGAAAGCGTCAACATGCAAAATTCGAAGATTGGCCA 1188
Qy      400 AsnSerLeuGlnAsnAlaThrHisLeuLeuLysGluSerLysMetLeuAsnLysThrPhe 419
Db      1189 TCACCTTACACATGCCATCAAGAAATTACGTAAGATCAAGTAATGATTGATGATTA 1248
Qy      420 GlyGluLysLeuLeuLeuHisTyrValAsnAlaAlaAsnValGluLeuLeuPheSer 439
Db      1249 GGTGCGTCATATTTTGGCAATTTTGTAGAAGCGAAGAAAGTGAATGGCAGCCTTCCT 1308
Qy      440 LysGlnValThrAspTrpGluLeuAsnGlnGlyPheAsnArgTyr 454
Db      1309 CAAACCGTTTCTGAATGGGAAGAGAACAAATATTGGAGTTGTAC 1353

RESULT 14
US-09-070-927A-188
; Sequence 188, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; Patrick J. Dillon
; Steven Barash
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 188:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4176 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 188:
US-09-070-927A-188

Alignment Scores:
Pred. No.: 1.57e-40 Length: 4176
Score: 402.00 Matches: 114
Percent Similarity: 44.81% Conservative: 63
Best Local Similarity: 28.86% Mismatches: 184
Query Match: 16.92% Indels: 34

```

```

DB:      10      7      Gaps:
US-10-098-602A-2 (1-454) x US-09-070-927A-188 (1-4176)
Qy      78 SerAspCysHisLeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTyrSerGluGly 97
Db      308 AGTGACATGATATTATACCTGACCTACTACATGCGATGATTTTCCGTGGGAAAGCGAC 367
Qy      98 -----AlaValLeuAlaLysSerAsnProHisAsnPheValThrSer 111
Db      368 CATGAAAGTTGCCGCTTAATTTGTGATATTACATCTCT-----GATGGT 415
Qy      112 GluProLeuPheCysSerProArgValLysLeuMetGlnGlnLysLeuArgLeuAlaAsn 131
Db      416 ACCCTTTTGGCGGAGATCCCGC-----GATAAT 445
Qy      132 LeuLysLeuLysGlyLeuPheAlaSerGluLeuGluPheAsnLeuPheAsn----- 148
Db      446 TTAACACGCGCTTAGCAGATATGAAGAAATTAGGCTTTACTCTTTTAACTTAGACCA 505
Qy      149 -----GluThrTyrLysSerAlaSerGlnLysHisTyrLysAsnLeuLysThrAlaGln 166
Db      506 GAACCAAGATTTTCTCTATTTAAATAGATGAGATGGTGAATATACCACTGACTGAAAT 565
Qy      167 ProHisHisGlnTrpMetAsnLysSerAlaSerSerGlyLysLeuThrPheMetArgSer 186
Db      566 GATAAAGTGGCTATTTGATTTTGCACCCGATTTAGTGAAATTTGTCGCGTGAT 625
Qy      187 ValArgAsnLysLeuGluAlaGlyLysLeuMetGluAlaThrHisProGluPheLeu 206
Db      626 ATTGTTTAGAGTTAGAAAGTTTAGGCTTTGAAGTGAAGCTTACATCATGAAGTAGCA 685
Qy      207 ProSerGlnHisGluLeuAsnPheValProAlaAspProLeuThrMetAlaAspArgHis 226
Db      686 CTGTGCAACATGAATAATGCTGATGTAGTTGATGCTGTGCAATATT 745
Qy      227 IleIleAlaLysHisGlyValArgGluMetAlaGluGlnSerGlyMetValAlaThrPhe 246
Db      746 CAACGTTCAATATTAGTTGTGAAGAACTATTGGCGCTAAACATGGCTTGCATGCACATTT 805
Qy      247 MetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeuGln 266
Db      806 ATGCCAAACCTTTGTTGGAATTAATGCTTCTGCTATGCACCTGCAATATGTCATTA--- 862
Qy      267 AspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMetSerThr 286
Db      863 ---TTTAATGAAGAGGTAATGTTTTTATGACGAGTACAGAGAGATGGCTTTAGACCAA 919
Qy      287 LeuAlaArgAsnTrpIleAlaGlyLeuLysTyrValProGluAlaThrTyrPhePhe 306
Db      920 ACTGGTATCATTTCTTAGTGGTTTATTAACATGCCCCGCTGCTATCTGCGGTATGT 979
Qy      307 AlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThrLysCys 326
Db      980 AACCCAAACAGTCACTCTTATAACGCTTGTGACAGGTTATGAAGCACCTGTTTATGTA 1039
Qy      327 CysTrpAlaLysAsnArgThrSerAlaPheArgLeuCysAsnSerLysSerGluGly 346
Db      1040 GCTTGGAGCGGAGAAATCGTTTCAACCACTCATTCGAGTTCTCGAATCTCGTGKTTATCG 1099
Qy      347 IleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsnProTyrLeuAlaPheSerAla 366
Db      1100 ACTCGTTTAGAATTAGTTCTGCTGCTCCTTACGCAATCCGTTATTAACTAGGCTGTC 1159
Qy      367 IleIleAlaAlaGlyLysSerGlyLysGluLysLeuGluLeuProProProAlaSer 386
Db      1160 TTATTACAAGCAGGTTTGGATGCTTTCGTAATGAACCTTACACCCAGCAGCTTGTAT 1219
Qy      387 GlyAsnValTyr-----AsnAspLysGlu-----LeuProGluPhePro 399
Db      1220 CGTAACATTTATGATGATGAGGAGAAAGCGTCAACATGCAAAATTCGAAGATTGGCCA 1279
Qy      400 AsnSerLeuGlnAsnAlaThrHisLeuLysGluSerLysMetLeuAsnLysThrPhe 419

```

```
Db 1280 TCAACCTTACCAATGCGCATCAAGAAATACGTAAGATCAAGTAATGATGCAATTA 1339
Qy 420 GlyGluLysLeuLeuHisTyrValAsnAlaAlaAsnValGluLeuAsnGluPheSer 439
Db 1340 GGTGCTCATATTTTGGCAATTTTGTAGAAGCAAGCAATGGAATGGGCGACGCTTCGCT 1399
Qy 440 LysGlnValThrAspTTPGluLeuAsnGlnGlyPheAsnArgTyr 454
Db 1400 CAACCCGTTCTGAATGGGAAAGAGAACAAATATTTTGGAGTTGTAC 1444
```

## RESULT 15

```
US-09-974-300-261
; Sequence 261, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085-500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 261
; LENGTH: 1008
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-261
```

## Alignment Scores:

```
Pred. No.: 9.5e-41 Length: 1008
Score: 395.50 Matches: 94
Percent Similarity: 51.25% Conservative: 50
Best Local Similarity: 33.45% Mismatches: 128
Query Match: 16.65% Indels: 9
DB: 10 Gaps: 2
```

US-10-098-602A-2 (1-454) x US-09-974-300-261 (1-1008)

```
Qy 181 GluThrPheMetArgSerValArgAsnLysLeuGluAlaGlyIleLeuMetGluAla 200
Db 171 GAAAACTGCCCGCGCATCGTACTTGAGCTCGAAGAAATGGGCTTTGAAATCGAAGCG 230
Qy 201 ThrHisProGluPheLeuProSerGlnHisGluLeuAsnPheValProAlaAspProLeu 220
Db 231 TCTCACCACGAAGTTGCACCTGGACAGCAGCAATCGACTTCAATACGCGCGAGCCATC 290
Qy 221 ThrMetAlaAspArgHisIleIleAlaLysHisGlyValArgGluMetAlaGluGlnSer 240
Db 291 CGCGCTTGGCATGACATCCAAACGTTCAAGCTCGTTGTTAAACAAATCGCGCGCAACAC 350
Qy 241 GlyMetValAlaThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHis 260
Db 351 GGCCTGACGCAACATTCATCCCAAAACCGTTGTTGCGCGTAAACGGATCAGGATGCAC 410
Qy 261 IleHisMetSerLeuGlnAspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAsp 280
Db 411 TGCATCTATCACTC-----TTCGAAACCGCGCCCAACGCTCTTTGACAAAGAGCCG 464
Qy 281 GluTyrGlyMetSerThrLeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyrValPro 300
Db 465 GATCTTCAGCTGAGTGAACCGCAAGCAATTCATCGCGGCATCGTCAAGCAGCAACA 524
Qy 301 GluAlaThrTyrPheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThr 320
Db 525 AGCTTTACAGCTGTGACAAACCCGACTGTCAACTCTTACAAAGCGCCTTTGTTCCAGGCTAT 584
```

```
Qy 321 PheAlaProThrLysCysCysTTPAlaIleAspAsnArgThrSerAlaPheArgLeuCys 340
Db 585 GAAGCAGCTTGTCTATGTGGCATGGCGCGCAAAACAGAACGCCGTTTAATCCGCATCCG 644
Qy 341 AsnSerLysSerGluGlyIleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsnPro 360
Db 645 GTTTCGCCGCGCATCAGCAGCGGTAGAAGTCAAGACGCTAGACCTTCTGCGCAACCCG 704
Qy 361 TyrLeuAlaPheSerAlaIleIleAlaAlaGlyIleSerGlyIleGluGluLysLeuGlu 380
Db 705 TACCTTGCTTGGCGTACTTCTTGCAGCAGGACTTGACGGCATCAAAAACAGCTGAA 764
Qy 381 LeuProProAlaSerGlyAsnValTyr-----AsnAspLys 393
Db 765 GCACCGGCTCCGATCGACCGCAACATCTATGTCTATGACAAAAGAGCGCTCGAAAAC 824
Qy 394 GluLeuProGluPheProAsnSerLeuGlnAsnAlaThrHisLeuLeuLysGluSerLys 413
Db 825 GGAATCGTCGACCTTCTCTGCGACACTCGCAGACGCTCTCGAAGAATTCAATCAAAACGAA 884
Qy 414 MetLeuAsnLysThrPheGlyGluLysLeuIleLeuHisTyrValAsnAlaAlaAsnVal 433
Db 885 GTCATGTCGTCAGGCCCTCGGCGGACCATCTATTGGAACACTTCGTCGAGCAAGAAATC 944
Qy 434 GluIleAsnGluPheSerLysGlnValThrAspTrpGluLeuAsnGlnGlyPheAsnArg 453
Db 945 GAATGGGATATGTTCCGCACACAGTCCATCCTTGGGAGCGGAGCAGTATATGTCTCAG 1004
Qy 454 Tyr 454
Db 1005 TAT 1007
```

Search completed: December 18, 2003, 07:57:15

Job time : 19019 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 17, 2003, 22:53:00 ; Search time 299 Seconds  
(without alignments)  
670.194 Million cell updates/sec

Title: US-10-098-602A-2

Perfect score: 2376

Sequence: 1 MTITYDLNLLIRNGKIDTV.....INFSKQVTDWELNQGNRY 454

Scoring table:

BLOSUM62	
Xgapop 10.0, Xgapext 0.5	
Ygapop 10.0, Ygapext 0.5	
Fgapop 6.0, Fgapext 7.0	
Delop 6.0, Delext 7.0	

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US10098602/runat\_17122003\_151238\_28076/app.query.fasta\_1.647  
-DB=Issued\_Patents\_NA -QPMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOFCU=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10098602 @CGN 1.1.56 @runat\_17122003\_151238\_28076 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEOQUERY -NEG SCORES=0 -WAIT -DSFBLA=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	799.5	33.6	4403765	3	US-09-103-840A-2
C 2	799.5	33.6	4411529	3	US-09-103-840A-1
C 3	481	20.2	1350	4	US-09-252-991A-1330
C 4	481	20.2	2538	4	US-09-252-991A-1285
C 5	481	20.2	2571	4	US-09-252-991A-1279
C 6	459	19.3	1347	4	US-09-252-991A-10926
C 7	459	19.3	1395	4	US-09-252-991A-10988
C 8	455	19.1	1395	4	US-09-252-991A-11477
C 9	452	19.0	1452	4	US-09-252-991A-11417
C 10	450	18.9	1470	4	US-09-252-991A-11262
C 11	449	18.9	4259	4	US-09-155-183-12
C 12	443	18.6	1425	4	US-09-252-991A-10987

C 13	432	18.2	13425	4	US-08-961-527-151	Sequence 151, App
C 14	409	17.2	1359	4	US-09-134-001C-1336	Sequence 1336, Ap
C 15	401.5	16.9	1359	4	US-09-252-991A-11454	Sequence 11454, A
C 16	398.5	16.8	1664976	4	US-08-916-421B-1	Sequence 1, Appli
C 17	393.5	16.6	1341	3	US-08-724-814-19	Sequence 19, Appl
C 18	389	16.4	1596	4	US-09-252-991A-10119	Sequence 10119, A
C 19	389	16.4	2400	4	US-09-252-991A-10051	Sequence 10051, A
C 20	360	15.2	3114	4	US-09-252-991A-225	Sequence 225, App
C 21	344.5	14.5	1185	4	US-09-252-991A-11261	Sequence 11261, A
C 22	344.5	14.5	1476	4	US-09-252-991A-2443	Sequence 2443, Ap
C 23	331.5	14.0	1443	4	US-09-328-352-2823	Sequence 2823, Ap
C 24	323.5	13.6	1488	4	US-09-252-991A-1197	Sequence 1197, App
C 25	311	13.1	576	4	US-09-252-991A-11196	Sequence 1196, A
C 26	303	12.8	1437	3	US-08-724-814-15	Sequence 15, Appl
C 27	303	12.8	4403765	3	US-09-103-840A-2	Sequence 2, Appli
C 28	303	12.8	4411529	3	US-09-103-840A-1	Sequence 1, Appli
C 29	293.5	12.4	906	4	US-09-252-991A-2239	Sequence 2239, Ap
C 30	287	12.1	1830121	4	US-09-557-884-1	Sequence 1, Appli
C 31	287	12.1	1830121	4	US-09-643-990A-1	Sequence 1, Appli
C 32	280.5	11.8	1416	4	US-09-252-991A-2530	Sequence 2530, Ap
C 33	280.5	11.8	1791	4	US-09-252-991A-2304	Sequence 2304, Ap
C 34	280.5	11.8	1986	4	US-09-252-991A-2389	Sequence 2389, Ap
C 35	272.5	11.5	756	4	US-09-252-991A-216	Sequence 216, App
C 36	267.5	11.3	1392	4	US-09-252-991A-9863	Sequence 9863, Ap
C 37	265.5	11.2	1446	4	US-09-252-991A-16240	Sequence 16240, A
C 38	265.5	11.2	1500	4	US-09-252-991A-16351	Sequence 16351, A
C 39	255.5	10.8	1470	4	US-09-328-352-237	Sequence 237, App
C 40	253.5	10.7	948	4	US-09-252-991A-16139	Sequence 16139, A
C 41	250.5	10.5	669	4	US-09-252-991A-11057	Sequence 11057, A
C 42	249	10.5	429	4	US-09-252-991A-11056	Sequence 11056, A
C 43	241.5	10.2	573	4	US-09-107-532A-959	Sequence 959, App
C 44	207.5	8.7	420	4	US-09-252-991A-11498	Sequence 11498, A
C 45	194.5	8.2	591	4	US-09-252-991A-208	Sequence 208, App

#### ALIGNMENTS

##### RESULT 1

US-09-103-840A-2/c

; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; TITLE OF INVENTION: TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; OTHER INFORMATION: CDC 1551  
; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
; OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Alignment Scores: 2.63e-78 Length: 4403765  
Pred. No.: 799.50 Matches: 176  
Score: 58.02% Conservatives: 88  
Best Local Similarity: 38.68% Mismatches: 180  
Query Match: 33.65% Indels: 11  
DB: 3 Gaps: 7

US-10-098-602A-2 (1-454) x US-09-103-840A-2 (1-4403765)

Qy	3	IleThrTyrAspGluLeuAanAsnLeuIleAargAenGlyLysIleAephThrValValLeu	22
Db		TTGGCGTGGACCGAGTTGGACGCGACTGGTGGCGCGGTGACGTGCGACACCGTCATCGTC	3167086
Qy	23	AlaCysValAspMetGlnGlyArgLeuMetGlyLysAargLeuThrGlyArgHisPheLeu	42
Db		CGCTTACCGACATGCAAGGCGCGCTGGCGGCAACCGATATACGGCGCCGCGATTTCGTC	3167026
Qy	43	--GlyLeuAspGlnLysLysIleSerThrPheValTyrAlaValThrIleGlu	61
Db		GAACATATGACACCGCGCGGTGCAGTGTGCAGTTATCTGTGGCCGTGGAGCTGCAC	3166966
Qy	62	GlyIleAlaGlyGlyTyrGluIleSerSerValAspThrGlyTyrSerAspCysHis	81
Db		CTGAACACGGTCCCGGCTATGCGATGGCCAGTTGGGACACCGGCTACGCGCATATGGTG	3166847
Qy	82	LeuCysAlaAspLeuAanSerLeuHisLeuLeuProTrpSerGluGlyAlaValLeuAla	101
Db		ATGACGGCGGAGTTGTCCACTCTCGGCTGATTCTCTGGGTACCGGGAACGGCGCTGGTG	3166784
Qy	102	IleSerAenProHisAanPheValThrSerGluProleuPheCysSerProArgValIle	121
Db		ATCGCCGAC---CTGGTCTGGCGCGACGGAGGAGTGCCTGCTCGCCGCGCAGCAATT	3166730
Qy	122	LeuMetGlnGlnIleGluArgLeuAlaAanLeuLysLeuLysGlyLeu	138
Db		CTGGCGGTGAGCTCGATCGG-----CTCAAGCGCGCGGACTGGTCCGCGATGTG	3166729
Qy	139	AlaSerGluLeuGluPheAanLeuPheAanGluThrTyrLysSerAlaSerGlnLysHis	158
Db		GCCACCGAGCTGGAGTTTCATGTTCTCGACCGCGGTATCCGACGCGATGGCGCAGCGG	3166678
Qy	159	TrpLysAanLeuLysThrAlaGlnProHisIleGlnTrpMetAenIleSerAlaSerSer	178
Db		TATCGGGGTGACCCCGCGCGAGCATCAACATCGACTACGCGATTTGGCGATCTCTCG	3166618
Qy	179	GlyIleGluThrPheMetArgSerValArgAanLysLeuGluGluAlaGlyIleLeuMet	198
Db		CGGATGGAGCGCTGTGTCGCGGACATCCGGTTGGTATGCGCGGTGGGTCTCGCATTC	3166558
Qy	199	GluAlaThrHisProGluPheLeuProSerGlnHisGluLeuAanPheValProAlaAsp	218
Db		GAGCGGTCAAAAGCGAATCAACATGGCGCAGCAGGAGATCGGGTTCTTCTACGACGAG	3166498
Qy	219	ProleuThrMetAlaAspArgHisIleAlaLysHisGlyValArgGluMetAlaGlu	238
Db		CGCTGTGTACCTCGCGCAACCATCGCATCTTACAAGAACGGCGCAAGGAAATCGCGCAG	3166438
Qy	239	GlnSerGlyMetValAlaThrPheMetAlaLysLeuSerSerThrAlaLeuGlyValAsnAla	258
Db		CAGCAGCGCAAGACCTTAAGTTTCATGGCGAATACGATCGGCTCCGCGA---GGTAAATAGC	3166378
Qy	259	CysHisIleHisMetSerLeuGlnAspAlaGluThrGluLysAanAlaPheTyrAspGln	278
Db		TGTCACATCATGTCTCGTCGCTGGCGACGATGGCTCCGCG---GNGTTTCCGACAGT	3166321
Qy	279	AsnAspGluTyrGlyMetSerThrLeuAlaAargAsnTrpIleAlaGlyLeuLeuLysTyr	298
Db		AACGGCGCGCAGCGCATGTGTCGTATGTCGAGTTCGCGAGTTCGTCGCGCGCATTTGGCCACG	3166264
Qy	299	ValProGluAlaThrTyrPhePheAlaSerTyrIleAanSerTyrLysArgLeuGlnPro	318
Db		TTGGCGGAAATTCAGCTGTGCTATGCGCGGACCCGACCCCTTAACCTCTTACAAGCGATTGGCGAT	3166204
Qy	319	LeuThrPheAlaProThrLysCysCysTrpAlaIleAspAanArgThrSerAlaPheArg	338
Db		AGCAGTTTCGCGCGACGGCGCTGGCTGGGTGGGGGTGACAACTGCACCTCGCCCTGCGG	3166144
Qy	339	LeuCysAanSerLysSerGluGlyIleAanValGluLeuArgIleGlyGlyAlaAspLeu	358
Db		GTGTT---GGCCACGGCGAAAACATCCGGGTGGAATGCCGGGTCCCGCGGTGATGTC	3166084
Qy	359	AsnProTyrLeuAlaPheSerAlaIleAlaAlaGlyIleSerGlnIleGluGluLys	378

Db	3166027	AACACGATACCTGGCGGTGGCGGCTCTCATTTGCTGGAGGTTGTACGGTATCGAGCGGGC	3165968
Qy	379	LeuGluLeuProProAlaSerGlyAsnValTyrAsnAspLysGluLeuProGluPhe	398
Db	3165967	CTTACGTGCGCCGAGCCCTGTGTGCGCAACGCCCTACCAAGGCCCGCATGTGCNACGGCTG	3165908
Qy	399	ProAsnSerLeuGlnAsnAlaThrHisLeuLeuLysGluSerLysMetLeuAsnLysThr	418
Db	3165907	CCGGTTACGCTGCGCCGAGCGCGGGTGCTGTTCCGAGGATTTGCGCTGTGTGCGCAGGCG	3165848
Qy	419	PhoGlyGluLysLeuLeuHisTyrValAsnAlaAlaAsnValGluLeuLeuGluPhe	438
Db	3165847	TTTCGGCGAGGATGTTGTCGCGCACTACCTCAACACGCGGTGTGGAGCTGGCGGCGGTTTC	3165788
Qy	439	SerLysGlnValThrAspTrpGluLeuAenGlnGlyPheAsnArg	453
Db	3165787	AACCGCGCGTACCAGATTGGGAGAGGATACGTGGATTTGAGCGC	3165743
RESULT 2			
US-09-103-840A-1/c			
; Sequence 1, Application US/09103840A			
; Patent No. 6294328			
; GENERAL INFORMATION:			
; APPLICANT: FLEISCHMAN, Robert D.			
; APPLICANT: WHITE, Owen R.			
; APPLICANT: FRASER, Claire M.			
; APPLICANT: VENTER, John C.			
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM			
; TITLE OF INVENTION: TUBERCULOSIS			
; FILE REFERENCE: 24366-20007.00			
; CURRENT APPLICATION NUMBER: US/09/103,840A			
; CURRENT FILING DATE: 1998-06-24			
; NUMBER OF SEQ ID NOS: 2			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 1			
; LENGTH: 4411529			
; TYPE: DNA			
; ORGANISM: Mycobacterium tuberculosis			
; OTHER INFORMATION: H3/Rv			
US-09-103-840A-1			
Alignment Scores:			
Pred. No.:	2,64e-78	Length:	4411529
Score:	799.50	Matches:	176
Percent Similarity:	58.02%	Conservative:	88
Best Local Similarity:	38.68%	Mismatches:	180
Query Match:	33.65%	Indels:	11
DB:	3	Gaps:	7
US-10-098-602A-2 (1-454) x US-09-103-840A-1 (1-4411529)			
Qy	3	IleThrTyrAspGluLeuAenAenLeuIleArgAsnGlyLysIleAspThrValValLeu	22
Db	3172973	TTGCGTGGACCGAGTTGGAGCGACTGGTCGCGCGCGGTGACGTCGACACCGTCATCGTC	3172914
Qy	23	AlaCysValAspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyVargHisPheLeu	42
Db	3172913	GCCTTCACCGACATGACAGGCGCGGTGGCGGCAACCGATATCGGGCGCGCATTTTCGT	3172854
Qy	43	---GlyLeuAspGlnLysLysIleSerIleSerThrPheValTyrAlaValThrIleGlu	61
Db	3172853	GACGACATACCGACCGCGCGGTGCGAGTGCTCGAGTTATCTGCTGCGCGCTGGACGTCGAC	3172794
Qy	62	GlyIleAlaGlyGlyGlyTyrGluIleSerSerValAspThrGlyTyrSerAspCysHis	81
Db	3172793	CTGAACACGGTGGCGGTATGCGATGGCGGCGAGTTGGGACACCGGTCACGCGATATGGTG	3172734
Qy	82	LeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTyrSerCluclyAlaValLeuAla	101
Db	3172733	ATGACCGCGGACTTGTCACACTTCGCGGCTGATTCTTGCTACCGGACACGGCGCTGGTG	3172674
Qy	102	IleSerAsnProHisAsnPheValThrSerGluProLeuPheCysSerProArgValIle	121

```
Db 3172673 ATCCCGGAC---CTGGTCTGGCGGACGCGCAGCGAGTCTCGCGCTCGCGCGAGCATT 3172617
Qy 122 LeuMetGlnGlnIleGluAtrGluAlaAsnLeuLysLeuLysGlyLeu-----Phe 138
Db 3172616 CTGGCGGTGAGCTGATCGTGG-----CTCAAGCGCGCGAGCTGTGCGCGATGTG 3172566
Qy 139 AlaSerGluLeuGluPheAsnLeuPheAsnGluThrTyrlYsSerAlaSerGlnLysHis 158
Db 3172565 GCCACCGAGCTGGAGTTTCATCGTGTTCGACCGCGTATCCCGAGCGTGGCGCCAGCGGG 3172506
Qy 159 TrpLysAsnLeuLysThrAlaGlnProHisHisGlnTrpMetAsnIleSerAlaSerSer 178
Db 3172505 TATCGCGGCTGACCGCGCGCAGCGACTACACATCGACTACCGGATATTCGCATCTCTCG 3172446
Qy 179 GlyIleGluThrPheMetArgSerValArgAsnLysLeuGluAlaGlyIleLeuMet 198
Db 3172445 CGGATGAGCGGCTGCTCGCGACATCCGGTGGGTATGGCGGTGCGGTCTCGGATTC 3172386
Qy 199 GluAlaThrHisProGluPheLeuProSerGlnHisGluLeuAsnPheValProAlaAsp 218
Db 3172385 GAGCGGTCAAGCGGATGCAACATGGCGCCAGCAGGAGATCGGTTTCGTTACGACGAG 3172326
Qy 219 ProLeuThrMetAlaAspArgHisIleIleAlaLysHisGlyValArgGluMetAlaGlu 238
Db 3172325 GCGCTGGTCACTCTCGCAACCATCGCATCTACAAGACGCGCGCAAGGAAATCGCGAC 3172266
Qy 239 GlnSerGlyMetValAlaThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAla 258
Db 3172265 CAGCAGCGCAAGAGCCTAACGTTTCATGGCGAAATACGATGAACCGGAA---GGTAATAGC 3172209
Qy 259 CysHisIleHisMetSerLeuGlnAspAlaGluThrGluLysAsnAlaPheTyrlYsAspGln 278
Db 3172208 TGTCAATCCATGTCCTCGCTCGCGGCGACGATGCTCGCG---GTGTTTCGCGACAGT 3172152
Qy 279 AsnAspGluThrGlyMetSerThrLeuAlaArgAsnTrpIleAlaGlyLeuLysTyrl 298
Db 3172151 AACGGCGCGCAGCGCATGTCGTCGATGTTCCGACAGTTCGTCGCGCGCAGTGGCCACG 3172092
Qy 299 ValProGluAlaThrTyrlPhePheAlaSerTyrlYsAsnSerTyrlYsAsnGlnPro 318
Db 3172091 TTGCGGAAATTCAGCTGTGTATGCGCGCGACCATTAACCTCAAGCGATTTGCCGAT 3172032
Qy 319 LeuThrPheAlaProThrLysCysTrpAlaIleAspAsnArgThrSerAlaPheArg 338
Db 3172031 AGCAGTTTCGCGCGCAGCGCGCTGCTTGGGGCTGGACATCGCATCTGCGCGCTGCGG 3171972
Qy 339 LeuCysAsnSerLysSerGluGlyIleAsnValGluLeuArgIleGlyGlyAlaAspLeu 358
Db 3171971 GTGGT---GGCCACGGGCAAAACATCCGGTTCGAATCCCGGTTCCCGCGGTGATGTC 3171915
Qy 359 AsnProTyrlLeuAlaPheSerAlaIleIleAlaGlyIleSerGlyIleGluLys 378
Db 3171914 AACCAGTACCTGGCGGTGGCGGCTCTCATTCCTGAGGGTTGTACGGTATCGACGGGGC 3171855
Qy 379 LeuGluLeuProProAlaSerGlyAsnValTyrlYsAsnAspLysGluLeuProGluPhe 398
Db 3171854 CTTAGCTGCCGAGCCCTGTGTCGCGACCGCTACCAAGCGCGGATGTCGACCGGCTG 3171795
Qy 399 ProAsnSerLeuGlnAsnAlaThrHisLeuLysGluSerLysMetLeuAsnLysThr 418
Db 3171794 CCGGTTCAGCTGGCGCGCGCGCTGTGTTCGAGGATTCGCGCTGGTGGCGGAGCGG 3171735
Qy 419 PheGlyLysLeuLeuLeuHisTyrlValAsnAlaAlaAsnValGluIleAsnGluPhe 438
Db 3171734 TTCGCGGAGGATGTGTTCGCGCATCTACTGAACACGCGGTGTGGAGCTGGCGCGCTTC 3171675
Qy 439 SerLysGlnValThrAspTrpGluLeuAsnGlnGlyPheAsnArg 453
Db 3171674 AACCGCGCGGTACCGATTTGGAGAGGATACGTGGATTTGAGCGC 3171630
```

RESULT 3

US-09-252-991A-1330

```
; Sequence 1330, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1330
; LENGTH: 1350
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-1330

Alignment Scores:
Pred. No.: 1,35e-46 Length: 1350
Score: 481.00 Matches: 131
Percent Similarity: 50.45% Conservative: 91
Best Local Similarity: 29.77% Mismatches: 185
Query Match: 20.24% Indels: 34
DB: 4 Gaps: 10

US-10-098-602a-2 (1-454) x US-09-252-991A-1330 (1-1350)
Qy 22 LeuAlaCysVal-----AspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArg 39
Db 61 ATCGAATGCTGTCACCCCGACCTCAACGGGGTGC CGCGCGCAAGGTGATGACCGCGCAG 120
Qy 40 HisPheLeuGlyLeuAspGlnLysLysIleSerIleSerThrPheValThrAlaValThr 59
Db 121 GGCTTC-----CTCAGCGCGCGCGCTCGCATCTGCGCGT-----GGCGTGGCTG 165
Qy 60 IleGluGlyIleAlaGlyGlyGlyTyrlYsSerValAspThrGlyTyrlYsSerAsp 79
Db 166 CTGCAATGATCATGGCGGCTTACCCCGCGCGGCTTCTATGGCAGCGGATGGCGAC 225
Qy 80 CysHisLeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTrpSerGluGlyAlaVal 99
Db 226 CTGCGCTGTGTCGCGCGCGCGACCGAGTCCATCGCTGCGCTGCGTGGAGC-----273
Qy 100 LeuAlaIleSerAsnProHisAsnPheValThrSerGlu-----112
Db 274 -----AACACTCGCGCGCTTTCGCATCTGCGATCGCAGGAACTGGATGGCAGC 324
Qy 113 ProLeuPheCysSerProArgValIleLeuMetGlnGlnIleGluArgLeuAlaAsnLeu 132
Db 325 CCGTCCGGGTGTCCACCCCGCGCTGCTCAACGAGTGTGGCGCTACGCGCCCGCCAC 384
Qy 133 LysLeuLysGlyLeuPheAlaSerGluLeuGluPheAsnLeuPhe-----AsnGluThr 150
Db 385 GGCTGGCAACCGGTGGTGGCCACCGAATCGAGTCTTCGCTTCGCGCCCGAACACCGAT 444
Qy 151 TyrlYs-SerAlaSerGlnLysHisTrpLysAsnLeuLysThrAlaGlnProHisHisGln 170
Db 445 CCCAACGAGCGGTTCAGGCGCGCGCTGGCGCGCGATGGCGCGCGAGCTG-GGCTATTC 503
Qy 170 nTrpMetAsnIleSerAlaSerSerGlyIleGluThrPheMetArgSerValArgAsnLys 190
Db 504 GGCGCTTCAGCGGTGTTCACCAACGCGCTGCGCGCGCTTCTTCGAGGACGCTTACCGCGT 563
Qy 190 sLeuGluGluAlaGlyIleLeuMetGluAlaThrHisProGluPheLeuProSerGlnHis 210
Db 564 CATGGATCGCTGGCGCTGTGTCGCGACACCTTCATCGACGAGATGGCGACCGAGT 623
Qy 210 sGluLeuAsnPheValProAlaAspProLeuThrMetAlaAspArgHisIleIleAlaLys 230
Db 624 CGAGATCAACTCTCTGTCATGGCGACCGCGTCTGCTGCGCGCGACGACGACCTTCCTGTTCAA 683
```





;; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

;; FILE REFERENCE: 107196.136  
 ;; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ;; CURRENT FILING DATE: 1999-02-18  
 ;; PRIOR APPLICATION NUMBER: US 60/074,788  
 ;; PRIOR FILING DATE: 1998-02-18  
 ;; PRIOR APPLICATION NUMBER: US 60/094,190  
 ;; PRIOR FILING DATE: 1998-07-27  
 ;; NUMBER OF SEQ ID NOS: 33142  
 ;; SEQ ID NO 10926  
 ;; LENGTH: 1347  
 ;; TYPE: DNA  
 ;; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-10926

Alignment Scores:  
 Pred. No.: 5,2e-44 Length: 1347  
 Score: 459.00 Matches: 126  
 Percent Similarity: 46.95% Conservative: 82  
 Best Local Similarity: 28.44% Mismatches: 213  
 Query Match: 19.32% Indels: 22  
 DB: Gaps: 10

US-10-098-602a-2 (1-454) x US-09-252-991A-10926 (1-1347)

QY 16 LysIleAspThrValValLeuAlaCysValAspMetGlnGlyArgLeuMetGlyLysArg 35  
 Db 29 GAGTCCAGTTCGTGACCTCTTATTCGACATGAAATGGCGTGTTCGGGCAAGCG 88  
 QY 36 LeuThrGlyArgHisPheLeuGlyLeuAspGlnLysLysIleSerIleSerThrPheVal 55  
 Db 89 ATCGAACGCAACAGTCTGAACAAAGTCTTCGAAAAGGTATCAATCTCCCGCTTCCCTG 148  
 QY 56 TyrAlaValThrIleGluGly-----IleAlaGlyGlyTyrGluIleSer 71  
 Db 149 TTGCGCTCGATATCACCGCTCCACCGTAGAGAGTACCGGCTGGGTCTCGACATCGCG 208  
 QY 72 SerValAspThrGlyTyrSerAspCysHisLeuCysAlaAspLeu--AsnSerLeuHis 90  
 Db 209 GAGCGCGAC-----CGATCTGCTACCGATCCCGCGACCTCTCC 250  
 QY 91 LeuLeuProThrSerGluGlyAlaValLeuAlaIleSerAsnProHisAsnPheValThr 110  
 Db 251 ATGGAACCTTGGCAGAACCGCGCGCACTCTGTGACCATGACGACGAACTGGAA 310  
 QY 111 SerGluProLeuPheCysSerProArgValIleLeuMetGlnGlnIleGluArgLeuAla 130  
 Db 311 GCGAGCCGTTCTTCGCCGACCCCGCAAGTCTCGCGCAGTGTGGCCAGGTTTCA 370  
 QY 131 AsnLeuLysLeuLysGlyLeuPheAlaSerGluLeuGluPheAsnLeuPheAsn--Glu 149  
 Db 371 GAGATGAAGTACCATCGCGCGCTTCGAGCTGGAGTTTACCTGTACGACGAGAG 430  
 QY 150 ThrTyrLysSerAlaSerGlnLysHisTrpLysAsnLeuLysThrAlaGlnProHisHis 169  
 Db 431 AAGGTGAACGGCGCGCGCGCGCTCGCTCGCGCATCTCCGCAAGGCTCG----- 484  
 QY 170 GlnTrpMetAsnIleSerAlaSerGlyIleGluThrPheMetArgSerValArgAsn 189  
 Db 485 CAGTCGCTGAGGTGTACTCCATCGACGACCTCGACGAATACTGTCAGTGCCTCCAGGAC 544  
 QY 190 LysLeuGluAla-----GlyIleLeuMetGluAlaThrHisProGluPheLeu 206  
 Db 545 ATCATGAGCGGCGCGCGCGCGCGCGCATCCCGCGCGACGCAATCTCGCCGCAATCCGCA 604  
 QY 207 ProSerGlnHisGluLeuAsnPheValProAla---AspProLeuThrMetAlaAspArg 225  
 Db 605 CCGGCGCAGTTCGAGTCAACCTCAACACGTCACGACGCGCTCAAGGCGCTGCGACAC 664  
 QY 226 HisIleIleAlaLysHisGlyValArgGluMetAlaGluLysSerGlyMetValAlaThr 245  
 Db 665 GCGGTGCTGCTCAAGCGCTGGTCAAGAACATCGCTACGACCCAGCAGATGACACCA 724

QY 246 PheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeu 265  
 Db 725 TTCATGGCCAGCCCTATCCGGGCGAGCGCGGAGGACTGCACGTGCATATCTCGCTG 784  
 QY 266 GlnAspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMetSer 285  
 Db 785 CTCGAC---AAGCATGGCAACAACATCTTACCAGGAGGATCCCGAG-----CAGAAC 835  
 QY 286 ThrLeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyrValProGluAlaThrTyrPhe 305  
 Db 836 GCCGATTGGCCCATCGCATCGCGTGTCTCGAGACCTCGCGGCTCCATGGCCCTTC 895  
 QY 306 PheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThrLys 325  
 Db 896 CTCCTGCCCGAAGTCACTCTACCGCGCTTCGGTTCGACGTTCTACGTGCCGAACGCG 955  
 QY 326 CysCysTrpAlaIleAspAsnArgThrSerAlaPheArgLeuCysAsnSerLysSerGlu 345  
 Db 956 CCGAGCTGGGCGCTGGCAACCGCACCGCTGGCGCTGCCGCTGCCACCGGACCGCGGAC 1015  
 QY 346 GlyIleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsnProTyrLeuAlaPheSer 365  
 Db 1016 GCGGTACCGCTGGAACACCGCGTGGCGCGCGCGCAACCCCTACTGCTACTGGCA 1075  
 QY 366 AlaIleIleAlaAlaGlyIleSerGlyIleGluGluLysLeuGluLeuProProAla 385  
 Db 1076 GCGGTGCTGGCAGGCGTTTCATCAGCGGTGACCAACAGGTGCGAGCGCGCGCGATC 1135  
 QY 386 SerGlyAsnValTyrAsnAspLysGluLeuProGluPheProAsnSerLeuGlnAla 405  
 Db 1136 GAAGGCAACTCTCTACGACGATGAG---CCGAGCCTGCCGAAACAACCTGCGCGCGC 1192  
 QY 406 ThrHisLeuLeuLysGluSerLysMetLeuAsnLysThrPheGlyGlyLeuLeuLeu 425  
 Db 1193 CTGCGCGAGCTGGAGAAAGGAGATCATGCGGAAGTACATCGACCCGGAAGTACATCGAC 1252  
 QY 426 HisTyrValAsnAlaAlaAsnValGluIleAsnGluPheSerLysGlnValThrAspTrp 445  
 Db 1253 ATCTCTGCTGCTGCAAGAAAGCGAGCTGGAGGATTCGAGCACTCGATCTCCGACCTC 1312  
 QY 446 GluLeuAsn 448  
 Db 1313 GAGTACAAC 1321

# RESULT 7

US-09-252-991A-10988  
 ; Sequence 10988, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 10988  
 ; LENGTH: 1395  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-10988

Alignment Scores:  
 Pred. No.: 5,5e-44 Length: 1395  
 Score: 459.00 Matches: 126  
 Percent Similarity: 46.95% Conservative: 82  
 Best Local Similarity: 28.44% Mismatches: 213  
 Query Match: 19.32% Indels: 22  
 DB: Gaps: 10



QY 111 SerGluProLeuPheCysSerProArgValIleLeuMetGlnGlnIleGluArgLeuAla 130  
 Db 364 GGTGATCCCTTCTCCGACACCCCGGAGTCTCCGCGAGGTGGCCAGGTTCCACC 423  
 QY 131 AsnLeuLysLeuLysGlyLeuPheAlaSerGluLeuGluPheAsnLeuPheAsn---Glu 149  
 Db 424 GAGATGAACTGACCATCGCGCGCTTCGAGTCTTACCTGATCGACGAGGAG 483  
 QY 150 ThrTyLysSerAlaSerGlnLysHisTrpLysAsnLeuLysThrAlaGlnProHisHis 169  
 Db 484 AAGTGAACGGCGCGCGAGCGCGCTCGCCGATCTCGGCAAGCGTCCG----- 537  
 QY 170 GlnTrpMetAsnIleSerAlaSerGlyIleGluThrPheMetArgSerValArgAsn 189  
 Db 538 CAGTCGGTCAAGTCTCATCGACGACCTCGCAATACGTCGAGTGGCTTCCAGGAC 597  
 QY 190 LysLeuGluGluAla-----GlyIleLeuMetGluAlaThrHisProGluPheLeu 206  
 Db 598 ATCATCGACGGCGCGCGCCAGGCGCATCCGCGCGACGCGCATCTCGCGAATCCGCA 657  
 QY 207 ProSerGlnHisGluLeuPhe-----ValProAlaAspProLeuThrMetAlaAspArg 225  
 Db 658 CGGGCGAGTTCAGGTCAACTGCATCATCGTCCGCGACCGCATGATGAAGCGCTCGACTAT 717  
 QY 226 HisIleAlaLysHisGlyValArgGluMetAlaGluLysSerGlyMetValAlaThr 245  
 Db 718 GGGTGTCTCAAGCGCTGATCAAGATCGCTCAGCACCAGCAGATGACACACC 777  
 QY 246 PheMetAlaLysLeuSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeu 265  
 Db 778 TTATCGCCCAAGCCCTATCGGGCCAGGCGGAGGACGAGTGCATCATCATCTCGCTG 837  
 QY 266 GlnAspAlaGluThrGlnLysAsnAlaPheTyAspGlnAsnAspGluTyGlyMetSer 285  
 Db 838 CTCGAC---AAGCACGGCAACACATCTTCCACGCGAGGATCCCGAG-----CAGAAC 888  
 QY 286 ThrLeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyValProGluAlaThrTyPhe 305  
 Db 889 CGGCGATTCGGCATCGCATCGCGCGGCTCGAGACCTCGCGCGCTCCATGGCGCTTC 948  
 QY 306 PheAlaSerTyLysAsnSerTyLysArgLeuGlnProLeuThrPheAlaProThrLys 325  
 Db 949 CTCTGCCGCAACCTCACTTACCGCGCTTCGGTTCGAGTCTACGTGCGGACGCG 1008  
 QY 326 CysCysTrpAlaIleAspAsnArgThrSerAlaPheArgLeuCysAsnSerLysSerGlu 345  
 Db 1009 CCGAGTGGGGCTGGACACCGCATCGTGGCGCTCGCGCTGCGCCACCGCGCGGAC 1068  
 QY 346 GlyIleAsnValGluLeuArgIleGlyAlaAspLeuAsnProTyLysLeuAlaPheSer 365  
 Db 1069 GCGGTACGCTGGAACATCGCTGCGCGCGCGCGAGCGCAACCCCTACCTCTGCTGGCGC 1128  
 QY 366 AlaIleAlaAlaGlyIleSerGlyIleGluGluLysLeuGluLeuProProAla 385  
 Db 1129 TCGGTCTTCGCGGGTGACACCGGCTGACCAACAGTTCGAGCGGGTGCGCGGATC 1188  
 QY 386 SerGlyAsnValTyLysAsnLysGluLeuProGluPheProAsnSerLeuGlnAsnAla 405  
 Db 1189 GAAGGCAACTCTTACGAGCAGGTTGGAG---CCGAGCGCTGCGGAACACCTGCGCGAGCGC 1245  
 QY 406 ThrHisLeuLeuLysGluSerLysMetLeuAsnLysThrPheGlyGluLysLeuLeu 425  
 Db 1246 CTCGGCAACTGGACGACGAGATCTCGGCGAAGTACATCATCGAAGTATCATCGAC 1305  
 QY 426 HisTyValAsnAlaAlaAsnValGluIleAsnGluPheSerLysGlnValThrAspTrp 445  
 Db 1306 ATCTTGTGCTGCAAGGAACGAGCTCGAGGAGTTCGAGTACTCGATCTCCGACCTC 1365  
 QY 446 GluLeuAsn 448  
 Db 1366 GAGTACAAAC 1374

## RESULT 9

US-09-252-991A-11417/c  
 ; Sequence 11417, Application US/09252991A  
 ; Patent No. 6551795

## GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 11417  
 ; LENGTH: 1452  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-11417

## Alignment Scores:

Pred. No.: 3,91e-43 Length: 1452  
 Score: 452.00 Matches: 122  
 Percent Similarity: 47.58% Conservative: 84  
 Best Local Similarity: 28.18% Mismatches: 205  
 Query Match: 19.02% Indels: 22  
 DB: 4 Gaps: 10

US-10-098-602A-2 (1-454) x US-09-252-991A-11417 (1-1452)

QY 26 AspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHisPheLeuGlyLeuAsp 45  
 Db 1449 GACATGAATGGCGTCTCGCGCAAGCAAGTCAAGCAACAGTCTGCCCAAGTCTTC 1390  
 QY 46 GlnLysLysIleSerIleSerThrPheValTyAlaValThrIleGluGly----- 62  
 Db 1389 GAAAGAGTATCAACTTCCCGTTCCTGTTGCGCTCGATATCACGGCTCCACCGTA 1330  
 QY 63 ----IleAlaGlyGlyTyArgLysSerSerValAspThrGlyTySerAspCysHis 81  
 Db 1329 GAGAGTACCGCGCTGGTCTCGACATCGCGCAGCGCGAC-----CGG 1288  
 QY 82 LeuCysAlaAspLeu---AsnSerLeuHisLeuLeuProTrpSerGluGlyAlaValLeu 100  
 Db 1287 ATCTGTACCCGATCCCGGACACCTCTCCATGGAACCTTGCAGAAAGCGCGCGCG 1228  
 QY 101 AlaIleSerAsnProHisAsnPheValThrSerGluProLeuPheCysSerProArgVal 120  
 Db 1227 CAACGTCTGATGACCATGCTGAGTGGAGGTGATCCCTTCTTCGCGACCCCGCGAA 1168  
 QY 121 IleLeuMetGlnGlnIleGluArgLeuAlaAsnLeuLysLeuLysGlyLeuPheAlaSer 140  
 Db 1167 GTCTCTCGCGCAGGTGGTGGCCAGGTTACCGAGATGGAACCTGACCATCTCGCGCTTC 1108  
 QY 141 GluLeuGluPheAsnLeuPheAsn---GluThrTyLysSerAlaSerGlnLysHisTrp 159  
 Db 1107 GAGTGGAGTCTTACTGATGACGAGAGAACGTAACCGCGCGCGCGCGCGCGCG 1048  
 QY 160 LysAsnLeuLysThrAlaGlnProHisHisGlnTrpMetAsnIleSerAlaSerSerGly 179  
 Db 1047 TCGCCGATCTCCGCAAGCGTCCG-----CAGTGGTGCAGGTGTACTCCATCGACGAC 994  
 QY 180 IleGluThrPheMetArgSerValArgAsnLysLeuGluAla-----GlyIle 196  
 Db 993 CTCGCAATACGTGAGTGGCTTCCAGGACATCATCGACGCGCGCGCGCGCGCGCATC 934  
 QY 197 LeuMetGluAlaThrHisProGluPheLeuProSerGlnHisGluLeuAsnPhe---Val 215  
 Db 933 CCGGCGGACGCGCATCTCGTCCCGCATCCGACCGCGCGCGCGCGCGCGCGCGCATC 874  
 QY 216 ProAlaAspProLeuThrMetAlaAspArgHisIleIleAlaLysHisGlyValArgGlu 235

[illegible]

## RESULT 10

```

US-09-252-991A-11262
; Sequence 11262; Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A

```

**Alignment Scores:**

Alignment scores:	
Pred. No.:	6.86e-43
Length:	1470

Score: 450.00

Percent Similarity:

```
QY 329 AlaIleAspAsnArgThrSerAlaPheArgLeuCysAsnSerLysSerGluGlyIleAsn 348
Db 1093 GCGGAAGAAGAACCGTACCGTCCGCTGCGCGTCCGCGATCCAGCCCGGAGAACCCCGG 1152
QY 349 ValGluLeuArgIleGlyGlyAlaAspLeuAsnProTyrLeuAlaPheSerAlaIleIle 368
Db 1153 GTGAGAACCGTCTCGCGCGCGCGAGCCCAACCCCTTACTTGGCGTGGCTGCCAGCCTG 1212
QY 369 AlaAlaGlyIleSerGlyIleGluGluLeuLysLeuLeuProProAlaSerGlyAsn 388
Db 1213 CTGTGCGGTACATCGCATGCTGCGAAGGATCAAGCCGAGCGCCCGAGGTCAAAGGTCC 1272
QY 389 ValTyrAsnAspLysGluLeuProGluPheProAsnSerLeuGlnAsnAlaThrHisLeu 408
Db 1273 GCGTATGAAGCGCGCAACCTG-----CGCTGCCCTGACGATCGAAGCGCGCTGGAGCGC 1329
QY 409 LeuLysGluSerLysMetLeuAsnLysThrPheGlyGlyLysLeuIleLeuHisTyrVal 428
Db 1330 ATGGAGAACTGCAAGCCCTCGAGCAGTATCTCGGCAGCAAGTTTCATCAGCGCTACGTC 1389
QY 429 AsnAlaAlaAsnValGluLeuLeuGluPheSerLysGlnValThrAspTrpGlu 446
Db 1390 GCGGTGAAGCGTCCGAGCAGCAGAGAAATTTCAAGCGGGTAATCAGCTCCTGGGAG 1443

RESULT 11
US-09-155-183-12
; Sequence 12, Application US/09155183
; Patent No. 6323011
; GENERAL INFORMATION:
; APPLICANT: Rhodes, Arjan
; APPLICANT: Rhodes, Michael J.C.
; APPLICANT: Gasson, Michael J.
; APPLICANT: Walton, Nicholas J.
; TITLE OF INVENTION: PRODUCTION OF VANILLIN.
; FILE REFERENCE: 20747/100
; CURRENT APPLICATION NUMBER: US/09/155,183
; EARLIER FILING DATE: 1999-05-03
; EARLIER APPLICATION NUMBER: PCT/GB97/00809
; EARLIER FILING DATE: 1997-03-24
; EARLIER APPLICATION NUMBER: GB96/06187
; EARLIER FILING DATE: 1996-03-23
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 4259
; TYPE: DNA
; ORGANISM: Pseudomonas fluorescens
US-09-155-183-12

Alignment Scores:
Pred. No.: 5,11e-42 Length: 4259
Score: 449.00 Matches: 128
Percent Similarity: 46.05% Conservative: 82
Best Local Similarity: 28.07% Mismatches: 224
Query Match: 18.90% Indels: 22
DB: 4 Gaps: 6

US-10-098-602A-2 (1-454) x US-09-155-183-12 (1-4259)
QY 1 MetThrIleThrTyrAspGluLeuAsnAsnLeuIleArgAsnGlyLysIleAspThrVal 20
Db 417 ATGAGTAACAACTCCAGCAGCTCACCAGTGTGGTTGAAGACCACAGATCAGAGATC 476
QY 21 ValLeuAlaCysValAspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHis 40
Db 477 GAATGCATGATTGGCGACTTGACCGGATCACCCGCGCAAGATCTCGCCCAACCAACAAG 536
QY 41 PheLeuGlyLeuAspGlnLysLysIleSerIleSerThrPheValTyrAlaValThrIle 60
Db 537 TTCATTGCCGAAAGAGCATGCGCCCTGCCGAGAGTGTGCTGTGAGACAGTACGCGGC 596
QY 61 GluGlyIleAlaGlyGlyGlyTyrGluIleSerSerValAspThrGlyTyrSerAspCys 80
```

```
Db 597 GACTATGTGGAAGACGACATCTAT---TACGAACTGCTCGACCCGCGCGACATCGACATG 653
QY 81 HisLeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTrpSerGluGlyAlaValLeu 100
Db 654 ATCTCCGCCCGGAGCAGAACCGGTGCTTCTCGTGCATGGCCATCGAGCCGCGCGG 713
QY 101 AlaIleSerAsnProHisAsnPheValThrSerGluProLeuPheCysSerProArgVal 120
Db 714 CAGGTGATTACGACACCTACGACAGGAGGCAACCCGATCGAGCTGCGCCAGCAAC 773
QY 121 IleLeuMetGlnGlnIleGluArgLeuAlaAsnLeuLysLeuLysGlyLeuPheAlaSer 140
Db 774 GTCTCTCAAGAAAGTCTCTCAAACTATTCCGCAAGAGGCTGGCAGCCGATCTGCGCGCG 833
QY 141 GluLeuGluPheAsnLeuPheAsnGluThrTyrLysSerAlaSerGlnLysHisTyrLys 160
Db 834 GAAATGGAGTTCTTACTCTG-----ACCAAGCGCAGTGCAGCAGCCCGATTAC----- 878
QY 161 AsnLeuLysThrAlaGlnPro-----HisHisGln 170
Db 879 -----CCATTGCAACCGCGGTGCGGTTCGGGACGTCCTCGGAATCGGTGCCCAA 929
QY 171 TrpMetAsnIleSerAlaSerSerGlyIleGluThrPheMetArgSerValArgAsnLys 190
Db 930 TCGTTCTCTATCGAAGCGGCCAACAAATTCCGACCGCTGTTCCGAAGACGCTCTACGACTGG 989
QY 191 LeuGluGluAlaGlyIleLeuMetGluAlaThrHisProGluPheLeuProSerGlnHis 210
Db 990 TCGCAACTGCGAGGAGCTGGATCTCGATACGCTGATCCACAGAGCGGACGCGCGAGATG 1049
QY 211 GluLeuAsnPheValProAlaAspProLeuThrMetAlaAspArgHisIleIleAlaLys 230
Db 1050 GAAATCAACTTCCGTACGCGGACGCGGTGCTCCCTGGCGGACGACATCTGTGTGTCAAG 1109
QY 231 HisGlyValArgGluMetAlaGluGlnSerGlyMetValAlaThrPheMetAlaLysLeu 250
Db 1110 CGCACCATGCGCGAGCGCGCTCAAGCAGCAACGTTGGCGGCCGCTTCATGGCCAAGCGG 1169
QY 251 SerSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeuGlnAspAlaGluThr 270
Db 1170 ATGACCGCGGAGCCTTGGCAGCGCCATGCACCTGCACGAGAGCATCATCGATATCGAGACC 1229
QY 271 GluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMetSerThrLeuAlaArgAsn 290
Db 1230 GGCAAGAACGCTTTC-----TCCATGAAGAGCGGAGCATGAGCAGTGTGTCTCTCAAC 1283
QY 291 TrpIleAlaGlyLeuLysTyrValProGluAlaThrTyrPhePheAlaSerTyrIle 310
Db 1284 CACATCGCGCGCTCGAGAAATTCATCCTGAACTGCTGCGCGCTGTTGCGGCCCAACGTC 1343
QY 311 AsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThrLysCysCysTrpAlaIle 330
Db 1344 AAATCGTTCCGCGCTTCTTCTGCGGACACTTCGCGCGCGGTGAACGTCGAGTGGGCGGAA 1403
QY 331 AspAsnArgThrSerAlaPheArgLeuCysAsnSerLysSerGluGlyIleAsnValGlu 350
Db 1404 GAAATCCGTTACCTGGCGCTGCGGTGCGGATGCGGCGCTCAAAACCGCTGGGTGGAA 1463
QY 351 LeuArgIleGlyGlyAlaAspLeuAsnProTyrLeuAlaPheSerAlaIleIleAlaAla 370
Db 1464 AACCGCTCGCGGTGCGGAGCCCAACCGTACCTGGCGGATTCGCGGAGCCTGCTGTGTC 1523
QY 371 GlyIleSerGlyIleGluGluLysLeuGluLeuProProAlaSerGlyAsnValTyr 390
Db 1524 GGCTACATCGGATGTCGAGGTATCAACCCAGCGCGCTGTGCTGGGTGCTGGTTAC 1583
QY 391 AsnAspLysGluLeuProGluPheProAsnSerLeuGlnAsnAlaThrHisLeuLys 410
Db 1584 GAGCGCGCAACCTG---CGTCTGCGCTGACCATCGAAGAGCTCTGGAACGCGATGGA 1640
QY 411 GluSerLysMetLeuAsnLysThrPheGlyGlyLysLeuIleLeuHisTyrValAsnAla 430
```

Db 1641 AACAGCAAGACCATCGAAGAAATACCTGGGTCAACAATTCATCATCTGCTAGCTCGCGGTC 1700  
 Qy 431 AlaAsnValGluIleAsnGluPheSerLysGlnValThrAspTrpGlu 446  
 Db 1701 AAGCGGCGGAGCATGAAACTTCAAGCGCGTGATCATGCTCATGGAA 1748

## RESULT 12

US-09-252-991A-10987/c  
 ; Sequence 10987, Application US/09252991A  
 ; Patent No. 6551795  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 10987  
 ; LENGTH: 1425  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-10987

## Alignment Scores:

Pred. No.: 4,34e-42 Length: 1425  
 Score: 443.00 Matches: 128  
 Percent Similarity: 45.92% Conservative: 80  
 Best Local Similarity: 28.26% Mismatches: 219  
 Query Match: 18.64% Indels: 26  
 DB: 4 Gaps: 7

US-10-098-602A-2 (1-454) x US-09-252-991A-10987 (1-1425)

Qy 6 AspGluLeuAsnLeuIleArgAsnGlyLysIleAspThrValValLeuAlaCysVal 25  
 Db 1422 GACAGCTGACCATCTGGTGAAGGACGAGGACGAGTACCGAGTGGATGCTGATTC 1363  
 Qy 26 AspMetGlnArgLeuMetGlyLysArgLeuThrGlyArgHisPheLeuGlyLeuAsp 45  
 Db 1362 GATCTGACAGCATCGCGCGGCAAGATCTCGCGGACCAACAATTCATCGCGGAAAG 1303  
 Qy 46 GlnLysLysIleSerIleSerThrPheValThrAlaValThrIleGluGlyIleAlaGly 65  
 Db 1302 GGCATGCGCTGCGGAGAGCGTCTGCTGAGACCGCTACCGCGGACTAGTGGAAAG 1243  
 Qy 66 GlyGlyTyrGluIleSerSerValAspThrGlyTyrSerAspCysHisLeuCysAlaAsp 85  
 Db 1242 GACATTAC---TACGACCTGCTCGACCGCGCGACATCGATATGGTCTGCGCGCGGAC 1186  
 Qy 86 LeuAsnSerLeuHisLeuLeuProTrpSer---GluGlyAlaValLeuAlaIleSerAsn 104  
 Db 1185 GAGAATCGCGTTCCTCTGCTCCCTGGCCATCGAGCCACCGCGGATGGTATCCAGC 1126  
 Qy 105 ProHisLeuPheValThrSerGluProLeuPheCysSerProArgValIleLeuMetGln 124  
 Db 1125 ACCTTCGCAAGCTC---GGCAATCCCATCGAGTGTGCGCGCAACATCTCTCAAGCGC 1069  
 Qy 125 GlnIleGluArgLeuAlaAsnLeuLysLeuLysGlyLeuPheAlaSerGluLeuGluPhe 144  
 Db 1068 GTGTGAGATGATGCGATAGGGCTGGCGCGGATCGTGGCGCAGAGATGGATTC 1009  
 Qy 145 AsnLeuPheAsnGluThr-----TyrLysSer 153  
 Db 1008 TACCTGACCAAGCGCAGCAGCCCGACTATCCGTTGAGGCGCGCGTGGCGCGCTCC 949  
 Qy 154 AlaSerGlnLysHisIleTrpLysAsnLeuLysThrAlaGlnProHisHisGlnTrpMetAsn 173  
 Db 948 GGACGCCAG-----GAAACCGGCGCTCAGTCTGCTTCTCC 916

Qy 174 IleSerAlaSerSerGlyIleGluThrPheMetArgSerValArgAsnLysLeuGluGlu 193  
 Db 915 ATCGAGCGCCCAACAGATTTCGAGGACATGTACGACTGCTGCGAGGCC 856  
 Qy 194 AlaGlyIleLeuMetGluAlaThrHisProGluPheLeuProSerGlnHisGluLeuAsn 213  
 Db 855 CAGGGCCCTTGACCTGGATACGCTGATCCACGAGGAGGACCGCGGAGATGGAGATCAAC 796  
 Qy 214 PheValProAlaAspProLeuThrMetAlaAspArgHisIleIleAlaLysHisGlyVal 233  
 Db 795 TTGCTCATCGGACGCGCTTCGACCTGGCGCAGATCTCTGGTGTTCAGCGGACCATG 736  
 Qy 234 ArgGluMetAlaGluGlnSerGlyMetValAlaThrPheMetAlaLysLeuSerThr 253  
 Db 735 CGCGAGCGCGCTCAAGCACAACGTCGCGCGACCTTCATGGCCAAGCCGATGACCGGC 676  
 Qy 254 AlaLeuGlyAsnAlaCysHisIleHisMetSerLeuGlnAspAlaGluThrGluLysAsn 273  
 Db 675 GAGCGGGCAGCGCATGCACTCCACAGAGCATCTCGACGCTGAAGACCGGCAAGAAC 616  
 Qy 274 AlaPheTyrAspGlnAsnAspGluTyrGlyMetSerThrLeuAlaIleAsnTrpIleAla 293  
 Db 615 ATCTTC-----TCCAATGCCGAGCGGACCATGAGCGAGCTGTTCTGCACTATATCGGC 562  
 Qy 294 GlyLeuLeuLysTyrValProGluAlaThrTyrPhePheAlaSerTyrIleAsnSerTyr 313  
 Db 561 GGCTGCGAGAGTTCATCCGGAAGTCTCCCGCTGTTTCGGGCCCAACGTGACTCTGTC 502  
 Qy 314 LysArgLeuGlnProLeuThrPheAlaProThrLysCysCysTrpAlaIleAspAsnArg 333  
 Db 501 CGCGCTTCTCGCGCGATACCTCGCGCGCGGTGAACGTCGAGTGGGCGGAGAGAACCGT 442  
 Qy 334 ThrSerAlaPheArgLeuCysAsnSerLysSerGluGlyIleAsnValGluLeuArgIle 353  
 Db 441 ACCGTGCGCTGCGCGTGGCGGACTCCAGCCCGGAGAACCGCGGTCGAGAACCGCTCTC 382  
 Qy 354 GlyGlyAlaAspLeuAsnProTyrLeuAlaPheSerAlaIleIleAlaAlaGlyIleSer 373  
 Db 381 CGCGGCGCGCAGCCCAACCTTACTTGGCGCTGGCTGGCGCTGCTGTCGCGCTACATC 322  
 Qy 374 GlyIleGluLysLeuLeuLeuProProAlaSerGlyAsnValTyrAsnAspLys 393  
 Db 321 GGCATGTGCAAGGCATCAAGCGCGCGCCAGGTCAAAGGTGCGGCTATGAACGCGC 262  
 Qy 394 GluLeuProGluPheProAsnSerLeuGlnAsnAlaThrHisLeuLeuLysGluSerLys 413  
 Db 261 AACCTG---CGCTGCGCTGACGATCGAAGCGCGCTGGAGCGCATGGAGAACTGCAAG 205  
 Qy 414 MetLeuAsnLysThrPheGlyGluLysLeuIleLeuHisTyrValAsnAlaAlaAsnVal 433  
 Db 204 CCCTCGAGCAGTATCTCGGACCAAGTTCATCAGCGGCTACGTCGCGTGAACGCTGCG 145  
 Qy 434 GluIleAsnGluPheSerLysGlnValThrAspTrpGlu 446  
 Db 144 GAGCAGCAGAAATTTCAAGCGGGTAATCAGCTCTCTGGAG 106

## RESULT 13

US-08-961-527-151/c  
 ; Sequence 151, Application US/08961527  
 ; Patent No. 6420135  
 ; APPLICANT: Charles Kunsch  
 ; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
 ; NUMBER OF SEQUENCES: 391  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Human Genome Sciences, Inc.  
 ; STREET: 9410 Key West Avenue  
 ; CITY: Rockville  
 ; STATE: Maryland  
 ; COUNTRY: USA  
 ; ZIP: 20850  
 ; COMPUTER READABLE FORM:

**MEDIUM TYPE:** Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

**CURRENT APPLICATION DATA:**

APPLICATION NUMBER: US/08/961,527

**FILING DATE:**

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

**ATTORNEY/AGENT INFORMATION:**

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REGISTRATION NUMBER: 30,373  
REFERENCE/DOCKET NUMBER: PB340P1

**TELECOMMUNICATION INFORMATION:**

TELEPHONE: (301) 309-8504

TELEPHONE: (301) 309-8512  
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 151:

### SEQUENCE CHARACTERISTICS:

LENGTH: 13425 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

STRANDEDNESS: none  
TOPOLOGY: linear

3-08-961-527-151

**Alignment Scores:**

Assigned Scores: 13425  
Assigned No.: 3.33e-39  
Length:

Score: 432.00  
Matches: 128

Score:	432.00	Matches:	120
Percent Similarity:	44.80%	Conservative:	83

Percent Similarity:	44.60%	Conservative:	83
Best Local Similarity:	27.18%	Mismatches:	220

Best Local Similarity:	27.18%	Mismatches:	220
Very March:	18.18%	Indels:	40

every match:	10:10%	10:10%
3:	4	10
		Gang:
		10

3-10-098-602A-2 (1-454) x US-08-961-527-151 (1-13425)

[illegible]

; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 1336  
; LENGTH: 1359  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-1336

Alignment Scores:  
Pred. No.: 4e-38 Length: 1359  
Score: 409.00 Matches: 111  
Percent Similarity: 47.84% Conservative: 77  
Best Local Similarity: 28.24% Mismatches: 175  
Query Match: 17.21% Indels: 30  
DB: 4 Gaps: 7

US-10-098-602A-2 (1-454) x US-09-134-001C-1336 (1-1359)

Qy 78 SerAspCysHisLeuLeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTyrSer----- 95  
Db 220 TCAGATATGATATTACATCTGATTTAGATACCTGGGTATCTTCCCTTGGACTGCTGGA 279  
Qy 96 GluGlyAlaValLeuAlaHisSerAsnProHisAsnPhenValThrSerGluProLeuPhe 115  
Db 280 CAAGGAAAGTTGCAGCACTAATCTGTGATGATTTAAACAGATGGTACACCATTTGAA 339  
Qy 116 CysSerProArgValLeuMetGlnLeuMetGlnLeuArgLeuAlaAsnLeuLeuLeuLeu 135  
Db 340 GGTATCCACGAGTAACTTGNAGCGTGTATTAGAGAAATGGAAGATATGGCTTTACT 399  
Qy 136 GlyLeu-----PheAlaSerGluLeuGluPheAsnLeuPheAsnGluThrTyrLeuSerAla 154  
Db 400 GATTTTAATCTAGGCGCTGAACCAAGATTTTCTATTATAATAGACCAAAAA----- 453  
Qy 155 SerGlnLysHisTyrPheAsnLeuLysThrAlaGlnProHisHisGln----- 170  
Db 454 -----GGCGAACCTACATAGTAATTAACAGATGAT 483  
Qy 171 -----TrpMetAsnIleSerAlaSerSerGlyLeuGluThrPheMetArgSerValArg 188  
Db 484 GGTGGTATTTCGATTAGTTCCTACAGATTAGTGAAATTTGTCGCGGTGACATCGTT 543  
Qy 189 AsnLysLeuGluGluAlaGlyLeuMetGluAlaThrHisProGluPheLeuProSer 208  
Db 544 TTAGAATTAGAAGATATGGCTTTGACATTGAAGCAAGCCACCATGAAGTACGCCAGGT 603  
Qy 209 GlnHisGluLeuAsnPhenValProAlaSerProLeuThrMetAlaAspArgHisIle 228  
Db 604 CAACATGAATTTGACTTTAAATATGCAGATCCGTTACAGCATGTGATAATATCCAAACA 663  
Qy 229 AlaLysHisGlyValArgGluMetAlaGluGlnSerGlyMetValAlaThrPheMetAla 248  
Db 664 TTTAACTAGTTGTAAACAACTTGCAGTGAAGTAAATTTACATGCAACATTTATGCCA 723  
Qy 249 LysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeuGlnAspAla 268  
Db 724 AAACCATTTATTTGGTGTAAACGGTAGTGTATGCATCTCAACGTATCACTA-----TTT 777  
Qy 269 GluThrGluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMetSerThrLeuAla 288  
Db 778 AAAGGAAAGAGATGCGCTCTTTGATCCTGAAGGTGATTTTACAAATTTGACTGATACTGCA 837  
Qy 289 ArgAsnTrpIleAlaGlyLeuLysTyrValProGluAlaThrTyrPhePheAlaSer 308  
Db 838 TATCAATTTACAGCTGGTGTCTTTAAACCGCTAGAGGATTCACCTGCAGTATGTAATCCA 897  
Qy 309 TyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThrLysCysCysTyr 328  
Db 898 ATTGTCAACTCATATAAGCTTCTTACAGGTTACGAGTACGAGCAACCATGTTATTATTCATGG 957  
Qy 329 AlaIleAspAsnArgThrSerAlaPheArgLeuCysAsnSerLysSerGluGlyLeuAsn 348  
Db 958 AGTGGTAAACCCGTTCACTTTAGTACGTGTCTCAACATCTAGAGGTCTATCAACTCGT 1017

## RESULT 15

US-09-252-991A-11454  
; Sequence 11454, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 11454  
; LENGTH: 1359  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-11454

## Alignment Scores:

Pred. No.: 3.05e-37 Length: 1359  
Score: 401.50 Matches: 103  
Percent Similarity: 49.27% Conservative: 65  
Best Local Similarity: 30.21% Mismatches: 162  
Query Match: 16.90% Indels: 11  
DB: 4 Gaps: 7

US-10-098-602A-2 (1-454) x US-09-252-991A-11454 (1-1359)

Qy 113 ProLeuPheCysSerProArgValIleLeuMetGlnIleGluArgLeuAlaAsnLeu 132  
Db 2 CCCTTCTTCGCCAGCCCGCGAAGTCTCGCCAGGTGGTGGCCAGGTTTCACCGAGATG 61  
Qy 133 LysLeuLysGlyLeuPheAlaSerGluLeuGluPheAsnLeuPheAsn---GluThrTyr 151  
Db 62 GAACGTACCATCTGTCGCCGCTTCGAGCTGGAGTTCTACCTGATCGCAGGAGAAAGTG 121  
Qy 152 LysSerAlaSerGlnLysHisTyrPheAsnLeuLysThrAlaGlnProHisHisGlnTrp 171  
Db 122 AACGGCGCGCGAGCGCGCGCTCGCCGATCTCCGCAAGCGTCG-----CAGTCG 175  
Qy 172 MetAsnIleSerAlaSerSerGlyIleGluThrPheMetArgSerValArgAsnLysLeu 191  
Db 176 GTGAGGTGTACTCCATCGACGACCTCGACGAATACGTGCGATCGCTCCAGGACATCATC 235

Qy	192	GlulGluala-----GlylleuMetGlualaThrHisProgluPheLeuProSer	208
Db	236	GAGGGCGCGCGCCCAAGGCGATCCCGCGCGAGCCATCGTCGCCGAATCCGACCGCGG	295
Qy	209	GlnHisgluLeuAenPhe---ValProAlaAspProLeuThrMetAlaAspArgHisIle	227
Db	296	CAGTTCGAGGTCAACTGTCATCAGTCGCCGACCCGATGAAGCGCTGCGACTATCGGGTG	355
Qy	228	IleAlaIysHisGlyValArgGluMetAlaGluGlnSerGlyMetValAlaThrPheMet	247
Db	356	CTGCTCAAGCGCTGATCAAGAACATCGCTCATGACCAACGAGATGGACACCACTTCATG	415
Qy	248	AlaIysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeuGlnAsp	267
Db	416	GCCAAAGCCCTATCCGGGCCAGGAGGAAACGGACTGCACATACACATCTCGCTGCTCGAC	475
Qy	268	AlaGluThrGluLysAenAlaPheTyAspGlnAsnAspGluTyArgLysMetSerThrLeu	287
Db	476	---AAGCACGGCAACATCTTCACACGAGGATCCCGAG-----CAGAACGCGCGCA	526
Qy	288	AlaArgAsnTrpIleAlaGlyLeuLeuLysTyArgValProGluAlaThrTyPhePheAla	307
Db	527	TTGCGCCATCGATCGCGCGGTGCTCGAGACCCTGCGCGCCTCCATGGCGCTTCTCTGCG	586
Qy	308	SerTyriIeAenSerTyriLysArgLeuGlnProLeuThrPheAlaProThrLysCysCys	327
Db	587	CCGAACGTCAACTCTTACCCGCGCTTCGGTTCGCGATTCTACGTGCCGAACGCGCGAGC	646
Qy	328	TrpAlaIleAspAenArgThrSerAlaPheArgLeuCysAenSerLysSerGluGlyIle	347
Db	647	TGGGGCTGGACAACCGCACCGTGGCCCTGCGGTGTCGCCACCGGACCGCGGCGGTA	706
Qy	348	AsnValGluLeuArgIleGlyGlyAlaAspLeuAenProTyriLeuAlaPheSerAlaIle	367
Db	707	CGCCTGGAAATCATCGCTGCGCGCGCGCGACCAACCCCTACCTGCTGCTGGCTCGGTG	766
Qy	368	IleAlaAlaGlyIleSerGlyIleGluGluLysLeuGluLeuProProAlaSerGly	387
Db	767	CTTGCCGGGTGCACACGSGGTGACCAACAAGTCGAGCGGGGTGCCCGATCGAAGGC	826
Qy	388	AsnValTyriAenApyLysGluLeuProGluPheProAenSerLeuGlnAenAlaThrHis	407
Db	827	AACTCTTACGAGCAGTGTGGAG---CCGAGGCTGCCGAACAACCTGCGCGACGCCCTGCGC	883
Qy	408	LeuLeuLysGluSerLysMetLeuAenLysThrPheGlyGluLysLeuIleLeuHisTyri	427
Db	884	GAACCTGGACACGCGAGATCCTTGGCGAAGTACATCATCGAAGTACATCGCATCTTC	943
Qy	428	ValAsnAlaAlaAenValGluIleAenGlnPheSerLysGlnValThrAspTrpGluLeu	447
Db	944	GTCCGCTCGAAGGAACGAGCTGGAGGAGTTCGAGTACTCGATCTCCGACCTCGAGTAC	1003
Qy	448	Asn	448
Db	1004	AAC	1006

Search completed: December 18, 2003, 04:28:12  
Job time : 7167 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 17, 2003, 22:45:30 ; Search time 394 Seconds  
(without alignments)  
3110.520 Million cell updates/sec

Title: US-10-098-602A-2

Perfect score: 2376  
Sequence: 1 MTITYDELNNLIRNGKIDTV.....INBFSQVTDWELNQGFNRY 454

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries.

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgp2\_1/USPTO.spool/US10098602/runat.17122003.151235.28032/app.query.fasta\_1.647  
-DB=N\_Geneseq\_19Jun03 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LIST=45 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-MODE=LOCAL -OUTFMT=ptc -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-USER=US10098602 @CGN 1.1 312 @runat.17122003.151235.28032 -NCPUP=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq\_19Jun03.\*  
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*  
25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	799.5	33.6	4403765	22	AAI99683	Mycobacterium tube
C 2	799.5	33.6	4411529	22	AAI99682	Mycobacterium tube
C 3	614.5	25.9	1860	22	AAI68686	Pseudomonas sp KIE
C 4	614.5	25.9	11355	22	AAI68692	Pseudomonas putida
C 5	502.5	21.1	34088	23	AAI59566	Propionibacterium
C 6	475.5	20.0	5500	24	ABQ82091	Brevibacterium lac
C 7	470.5	19.8	1338	22	AAH67407	C glutamicum codin
C 8	470.5	19.8	1461	22	AAH71800	Cornebacterium gl
C 9	470.5	19.8	349980	22	AAH68531	C glutamicum codin
C 10	464	19.5	2365589	24	ABA90521	Genomic sequence o
C 11	453	19.1	349980	24	ABQ81848	Bifidobacterium lo
C 12	449	18.9	4259	18	AAV04201	Merged contigs pfi
C 13	445	18.7	1344	24	ABN68493	Streptococcus poly
C 14	445	18.7	2155561	24	ABN71527	Streptococcus poly
C 15	441	18.6	1344	24	ABN68494	Streptococcus poly
C 16	432.5	18.2	1373	22	APF60972	P. putida KT2440-a
C 17	432	18.2	1344	25	ABX06145	S. pneumoniae type
C 18	432	18.2	1347	23	AAV52284	Streptococcus pneu
C 19	432	18.2	13425	19	AAV52284	Streptococcus pneu
C 20	432	18.2	2162598	25	ABS56454	Streptococcus pneu
C 21	426	17.9	1348	25	AAD51270	Streptococcus pneu
C 22	424.5	17.9	1338	23	AAV52096	Staphylococcus aur
C 23	424.5	17.9	1341	23	AAV54603	Staphylococcus aur
C 24	424.5	17.9	1341	23	AAV55040	Staphylococcus aur
C 25	411	17.3	495269	24	ABQ67195	Listeria innocua c
C 26	411	17.3	3011208	24	ABQ69245	Listeria innocua D
C 27	410	17.3	3497	24	ABQ70835	Listeria monocytog
C 28	410	17.3	2944528	24	ABA03041	Listeria monocytog
C 29	409	17.2	1341	22	AAH53428	S. epidermidis ope
C 30	409	17.2	1359	24	ABN91873	Staphylococcus epi
C 31	402	16.9	1356	23	AAV52780	Enterococcus faeca
C 32	402	16.9	4176	20	AAI13125	Enterococcus faeca
C 33	402	16.9	4176	24	ABS98920	Enterococcus faeca
C 34	400.5	16.9	1329	10	AAV20117	Enterococcus faeca
C 35	397.5	16.7	1664976	19	AAV21209	DNA encoding pepti
C 36	395.5	16.6	1008	24	ABK72970	Methanococcus jann
C 37	393.5	16.6	1341	19	AAV29071	Bacillus lichenifo
C 38	393.5	16.6	38110	25	AAV29071	Mycobacterium tube
C 39	393.5	16.6	38110	25	AAV29071	Mycobacterium tube
C 40	389.5	16.4	265118	22	AAH41227	Ribonuclease P RNA
C 41	387	16.3	3666	22	AAH54179	Pyrococcus abyssi
C 42	351	14.8	989	19	AAZ96252	S. epidermidis gen
C 43	351	14.8	989	19	AAZ96252	S. pneumoniae deri
C 44	338.5	14.2	3861	22	AAH54760	Streptococcus pneu
C 45	330.5	13.9	1344	22	AAV61030	S. epidermidis gen
C 46						P. putida KT2440-a

ALIGNMENTS

RESULT 1  
AAI99683/C  
ID AAI99683 standard; DNA; 4403765 BP.  
XX  
AC AAI99683;  
XX  
DT 15-JAN-2002 (first entry)  
XX  
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.  
XX  
KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;  
XX  
XX variation; epidemiology; patient treatment; epidemic monitoring; ds.  
XX  
OS Mycobacterium tuberculosis.  
XX  
PN US6294328-B1.  
XX



XX Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;  
 KW variation; epidemiology; patient treatment; epidemic monitoring; ds.  
 XX Mycobacterium tuberculosis.  
 OS US6294328-B1.  
 XX PD 25-SEP-2001.  
 XX PF 24-JUN-1998; 98US-0103840.  
 XX PR 24-JUN-1998; 98US-0103840.  
 XX PA (GENO-) INST GENOMIC RES.  
 XX PI Fleischmann RD, White OR, Fraser CM, Venter JC;  
 XX WPI; 2001-647261/74.  
 DR Evaluating strain variation of Mycobacterium tuberculosis, comprises  
 PT determining the nucleotide sequence of the strain at positions in the  
 PT genome corresponding to positions where M. tuberculosis strains CDC  
 PT 1551 and H37Rv differ.  
 XX Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.  
 XX The invention relates to evaluating strain variation within and between  
 CC different populations of the tuberculosis bacterial pathogen,  
 CC Mycobacterium tuberculosis or related Mycobacterium by determining the  
 CC nucleotide sequence of the first strain at positions in the complete  
 CC sequence of the genome that correspond to positions that differ in the  
 CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and  
 CC H37Rv (AA199682). The method is useful for evaluating strain variation of  
 CC M. tuberculosis and has valuable application in the fields of  
 CC tuberculosis genetics, epidemiology, patient treatment and epidemic  
 CC monitoring.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from USPTO  
 CC at [seqdata.uspto.gov/sequence.html?docID=6294328B1](http://seqdata.uspto.gov/sequence.html?docID=6294328B1).  
 XX SQ Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 4.58e-71 Length: 4411529  
 Score: 799.50 Matches: 176  
 Percent Similarity: 58.02% Conservative: 88  
 Best Local Similarity: 38.88% Mismatches: 180  
 Query Match: 33.65% Indels: 11  
 DB: 22 Gaps: 7  
 US-10-098-602A-2 (1-454) x AA199682 (1-4411529)  
 Qy 3 IleThrTyrAspGluLeuAsnAsnLeuLeuArgAsnGlyIleAspThrValValLeu 22  
 Db TTGGCGTGGACGAGTGGAGCGACTGCTCGCGCGCGTACGTCGACCGTCATCGTC 3172914  
 Qy 23 AlaCysValAspMetGlnArgLeuMetGlyValArgLeuThrGlyArgHisPheLeu 42  
 Db CGCTTCACCGACATGCGAGGCGCGTGGCGGCAACCGGATTCGGCGCGGCATTCGTC 3172854  
 Qy 43 --GlyLeuAspGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLys 61  
 Db GACGACATAGCCACCGCGCGTGCAGTGTGCGAGTATCTGCTGGCGCGTGCAGTCGAC 3172853  
 Qy 62 GlyIleAlaGlyGlyTyrGluIleSerValAspThrGlyTyrSerAspCysHis 81  
 Db CTGAACACGCGCGCGTATGCGATGCGCCAGTGGGACACCGCGTACGGCGATGCGT 3172734  
 Qy 82 LeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTrpSerGluGlyAlaValLeuAla 101  
 Db ATGACGCGCGACTTGTCCACTCTCGGCTGATTCCTTGGTACCGGGAACGCGCTGCTG 3172674

Qy 102 IleSerAsnProHisAsnPheValThrSerGluProLeuPheCysSerProArgValIle 121  
 Db ATGCCCGAC--CTGGTCTGGCGCGCGAGCGAGCGAGTGGCGGCTCGCGCGAGCATT 3172617  
 Qy 122 LeuMetGlnGlnIleGluArgLeuAlaAsnLeuLysLeuLysGlyLeu-----Phe 138  
 Db CTGCGCGTCACTCGATCGG-----CTCAAGCGCGCGCGAGTGGTCCGCGATGTG 3172566  
 Qy 139 AlaSerGluLeuGluPheAsnLeuPheAsnGluThrTyrLysSerAlaSerGlnLysHis 158  
 Db GCCACCGAGCTGGAGTTCATGTCGACCGCGTATGCCAGGATCGGCGCGCGG 3172506  
 Qy 159 TrpLysAsnLeuLysThrAlaGlnProHisGlnTrpMetAsnIleSerAlaSerSer 178  
 Db TATCGCGGCTGACCCCGCGCGAGCTCAACATCGACTACGCGATATTGGCATCTCTCG 3172446  
 Qy 179 GlyIleGluThrPheMetArgSerValArgAsnLysLeuGluAlaGlyIleLeuMet 198  
 Db CGGATGAGCGCGTGGTGGCGGATCGGCTGGGTATGGCGGTCGCGGTCCTCGATTC 3172386  
 Qy 199 GluAlaThrHisProGluPheLeuProSerGlnHisGluLeuAsnPheValProAlaAsp 218  
 Db GAGCGGTCAAGCGGATTCGACATGGCGCAGCAGAGATCGGTTTCGTACGACGAG 3172326  
 Qy 219 ProLeuThrMetAlaAspArgHisIleAlaLysHisGlyValArgGluMetAlaGlu 238  
 Db GCGCTGTCACCTGCGCAACCATGCGATCTACAAGAACGCGGCAAGGAATCGCGAC 3172266  
 Qy 239 GlnSerGlyMetValAlaThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAla 258  
 Db CAGCAGCGCAAGAGCCCTAACGTTTCATGGCGAAATACGATGAACCGCAA--GGTAATAGC 3172209  
 Qy 259 CysHisIleHisMetSerLeuGlnAspAlaGluThrGluLysAsnAlaPheTyrAspGln 278  
 Db TGTCACTCATCTGCTCGCTGCGTGGCAGCGATGGTCCGCG--GTGTTGCCGACAGT 3172152  
 Qy 279 AsnAspGluTyrGlyMetSerThrLeuAlaAsnTrpIleAlaGlyLeuLeuLysTyr 298  
 Db AACGGGCGCGACGCGATGTCGTGATGTTCCGCGAGCTTCGTCGCGCGCCAGTTGCCACG 3172092  
 Qy 299 ValProGluAlaThrTyrPhePheAlaSerTyrIleAsnSerTyrLysArgLeuGlnPro 318  
 Db TTGCGCGAATTCAGCTGCTGCTATGCGCGCACCATTAACCTCTACAAGCGATTTGCCGAT 3172032  
 Qy 319 LeuThrPheAlaProThrLysCysCysTrpAlaIleAspAsnArgThrSerAlaPheArg 338  
 Db AGCAGTTTCGCGCGCGCGCGTGGTGGGCTGGAGCAATCGCACTCGCCCTCGCGG 3171972  
 Qy 339 LeuCysAsnSerLysSerGluGlyIleAsnValGluLeuArgIleGlyAlaAspLeu 358  
 Db GTGGTT--GGCCACGGGCAAAACATCCGGTTCGATGCGCGGTTCCGCGGCGGTGATGTC 3171915  
 Qy 359 AsnProTyrLeuAlaPheSerAlaIleAlaAlaGlyIleSerGlyIleGluGluLys 378  
 Db AACAGTACCTGGCGGTGGCGGCTCTCATTCGTGGAGGTTGTACGATATCGAGCGGCG 3171855  
 Qy 379 LeuGluLeuProProAlaSerGlyAsnValTyrAsnAspLysGluLeuProGluPhe 398  
 Db CTTTCAGTCCCGCGAGCGCTGTGTCGCGCAACCGCTTACCAGCGCGGATGCGAACCGGCTG 3171795  
 Qy 399 ProAsnSerLeuGlnAsnAlaThrHisLeuLysGluSerLysMetLeuAsnLysThr 418  
 Db CCGGTTACGCTGGCGCGCGCGGTCGTCGAGGATTCGCGCTGGTGGCGGAGCGG 3171735  
 Qy 419 PheGlyLysLysLeuLeuLeuHisTyrValAsnAlaAlaAsnValGluIleAsnGluPhe 438  
 Db TTGCGCGAGATGTGTGCGCGCACTACTGAACAAACGCGGCTGTGGAGCTGGCGGCTTC 3171675  
 Qy 439 SerLysGlnValThrPheTrpGluLeuAsnGlnGlyPheAsnArg 453  
 Db AACCGCGCGGTCAACCGATTGGGAGAGGATACGTGGATTGAGCGC 3171630

RESULT 3



```

Db 1414 ATCGAAATAAATACTCAAGCCGAAAGATCCAAATCTCGGTAAACGCTTAC-----AAGATA 1467
Qy 395 LeuProGluPheProAsn-----SerLeuGlnAsnAlaThrHisLeuLeuLys 410
Db 1468 TCGCGGAGCTCGCTCGCCACTCGCGCGTCATTTGGAGGAGCTGCGGCATTTTCGC 1527
Qy 411 GluSerLysMetLeuAsnLysThrPheGlyGluLysLeuLeuLeuHisTyrValAsnAla 430
Db 1528 GAAAGTGAATGGCGGAGTGATCTTCCCTAACGAGTTCGTTGAGCACTATGCCAGATG 1587
Qy 431 AlaAsnValGluLeuAsnGluPheSerLysGlnValThrAspTyrGluLeu 447
Db 1588 AAAGTTTGGAGATTAAACAGTCCCAATAGTTTCGTGTAATAATTTGGAGTTG 1638

RESULT 4
ID AA168692
XX AA168692 standard; DNA; 11355 BP.
AC AA168692;
XX
DT 17-JAN-2002 (first entry)
XX
DE Pseudomonas putida ipuA, B, C, D, E, F, G and H encoding DNA.
XX
KW L-alaninol; isopropylamine; iupI; ipuH; ofloxacin; ipuC; glutamylamide;
KW gamma-glutamylamide synthase; theanine; ipuA; ipuB; ipuD; ipuE; ipuF;
KW ipuG; ipuH; ds.
XX
OS Pseudomonas putida.
XX
FH Key Location/Qualifiers
FT CDS 1314..2340
FT FT /*tag= a
FT FT /product= "ipuA"
FT FT 2342..2680
FT FT /*tag= b
FT FT /product= "ipuB"
FT FT 2743..4122
FT FT /*tag= c
FT FT /product= "ipuC"
FT FT 4194..5354
FT FT /*tag= d
FT FT /product= "ipuD"
FT FT 5371..5565
FT FT /*tag= e
FT FT /product= "ipuE"
FT FT 5589..6476
FT FT /*tag= f
FT FT /product= "ipuF"
FT FT 6533..7963
FT FT /*tag= g
FT FT /product= "ipuG"
FT FT 8051..9574
FT FT /*tag= h
FT FT /product= "ipuH"
FT FT 9600..9698
FT FT /*tag= i
FT FT /note= "Unidentified partial ORF, no start or stop
FT FT codon given"
XX
PN WO200173038-A2.
XX
PD 04-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-EP03651.
XX
PR 31-MAR-2000; 2000EP-0106888.
XX
PA (LONZ ) LONZA AG.
XX
PI Leisinger T, Van Der Ploeg J, Kiener AM, De Azevedo Waesch SI;
PI Maire T;
XX

```

```

DR WPI; 2001-626266/72.
DR P-PSDB; AAG80036, AAG80037, AAG80038, AAG80039, AAG80040, AAG80041,
DR AAG80042, AAG80043, AAG80044.
XX
PT New microorganism for converting isopropylamine to L-alaninol, useful
PT as pharmaceutical intermediate especially ofloxacin, has the ipuH and I
XX genes for alaninol catabolism inactivated
XX
PS Claim 8; Fig 3A-N; 106pp; German.
XX
CC This invention describes novel microorganisms (A), or their enzymatic
CC extracts, able to convert isopropylamine (I) to L-alaninol (II) and which
CC have the ipuH and ipuI genes, encoding enzymes involved in metabolism of
CC (II), inactivated. (II) is an intermediate for pharmaceuticals,
CC especially ofloxacin. Also the enzyme encoded by the bacterial ipuC
CC (a gamma-glutamylamide synthase) is used for synthesis of gamma-
CC glutamylamides, specifically theanine. Inactivation of the ipuH and I
CC genes improves production of (II). Typically, a Pseudomonas in which both
CC ipuH and I genes were inactivated produced (II) at 8 mM after 60 hour
CC culture. For a similar strain in which only the ipuI had been inactivated
CC the maximum (II) concentration was about 7 mM, after 6 hour, falling to
CC practically zero after about 20 hour. This sequence encodes the
CC Pseudomonas putida ipuA, ipuB, ipuC, ipuD, ipuE, ipuF, ipuG and ipuH
CC protein products which are described in the method of the invention.
XX
SQ Sequence 11355 BP; 2508 A; 2720 C; 3264 G; 2862 T; 1 other;

Alignment Scores:
Pred. No.: 5,95e-55 Length: 11355
Score: 614.50 Matches: 159
Percent Similarity: 50.77% Conservative: 73
Best Local Similarity: 34.79% Mismatches: 194
Query Match: 25.86% Indels: 31
DB: 22 Gaps: 10

US-10-098-602a-2 (1-454) x AA168692 (1-11355)
Qy 7 GluLeuAsnAsnLeuLeuLeuArgAsnGlyLysValValValValLeuAlaCysValAsp 26
Db 2773 AAGGTCGAGATTTTATTGAGAAACAAATATCGACAGATTAGGCTGGGTGCGAGTCAT 2832
Qy 27 MetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHisPheLeu-----Gly 43
Db 2833 ATAGATGGAGTCTGGCGCGCTAAGCAAGTTGGCGCAGAGTATTCTTTGAATAAGCGGCT 2892
Qy 44 LeuAspGlnLysLysLysLysSerIleSerThrPheValTyrAlaValThrIleGluGlyIle 63
Db 2893 CTTGATGGGACGAGATCTCTAAACATTCTGTTGGATGGGATGTGGCCGATCACTTCGTC 2952
Qy 64 AlaGlyGlyGlyTyrGluIleSerSerValAspThrGlyTyrSerAspCysHisLeuCys 83
Db 2953 -----GACGTTTGGAGTTTACTGGCTGGGATTCGGGTATCCGGATATCGCCTTAATT 3006
Qy 84 AlaAspLeuAsnSerLeuHisLeuLeuProTyrSerGluGlyAlaValLeuAlaIleSer 103
Db 3007 CCTGATCTTTTCGACTCTTTTCGTTGACCTGGCAGGAAAAAACTGCTTCGCTCCTTTCG 3066
Qy 104 AsnProHisAsnPheValThrSerGluProLeuPheCysSerProArgValIleLeuMet 123
Db 3067 GATATCCAGCAT---CTAAATGGCGAGCCGCTGAATCTTTTCGCTAGGAACCTTCTCGCT 3123
Qy 124 GlnGlnIleGluArgLeuAlaAsnLeuLysGlyLeuPheAlaSerGluLeuGlu 143
Db 3124 AAAGCAATCGAAAAAGCAGACGAGCTTGGTTACAGTGCTACGCTGCTTACGAGTTTGAG 3183
Qy 144 PheAsnLeuPheAsnGluThrTyrLysSerAlaSerGlnLysHisIleTyrLysAsnLeuLys 163
Db 3184 TTTTACCTTCTTAACGACTCTATCGCGAGTATTTCGCGACATCAGTCGCGCAGCATTAAC 3243
Qy 164 ThrAlaGlnPro-----HisHisGlnTrpMetAsnIleSer 175
Db 3244 CCGGTGAAAAAGAGCGGCGCACTGCTCAGCATGCTCCATCAC-----TCT 3288

```



Qy 74 AspThrGlyTyr-----SerAspCysHisLeuCysAlaAspLeuAsnSer 88  
Db 18790 GAG---GGGTATGCCCGCGTCTAGCAGGCAGATGTGTGGCCACCCGATCCGCCAC 18734  
Qy 89 LeuHisLeuLeuProTTPSerGluGlyAlaValLeuAla-----IleSer 103  
Db 18733 TTCCAAGTTCTGCCATGCGTTCGGACCGACCGCAGGATGTTCTGGGATATCACC 18674  
Qy 104 AsnProHisAsnPheValThrSerGluProLeuPheCysSerProArgValIleLeuMet 123  
Db 18673 AATCTCTGACGGA-----ACCTCTCTGCA-----GCTGATCCTCGTACGCTCTCAAG 18626  
Qy 124 GlnGlnIleGluArgLeuAlaAsnLeuLysLeuLysGlyLeuPheAlaSerGluLeuGlu 143  
Db 18625 AGGACGATGGTCTGCGCGCAGAAATGGTTTACCTTCTATGTTCCACCGAGATTGAG 18566  
Qy 144 PheAsnLeuPheAsnGluThrTyrLysSerAlaSerGlnLysHisTyrLysLeuLys 163  
Db 18565 TTCTACCTGCTTAACGAT-----GAGCATCATCTCTGGGGAATCCCGC 18524  
Qy 164 ThrAlaGlnProHisHisGlnTrpMetAsnIleSerAlaSerSerGlyIleGluThrPhe 183  
Db 18523 GTGGCGCTGGACACCGTGGCTACTTCGACCACACCTTGGGAGCGGCACCGACTTC 18464  
Qy 184 MetArgSerValArgAsnLysLeuGluAlaGlyIleLeuMetGluAlaThrHisPro 203  
Db 18463 CGTCGACGCGCATCAACGCTCGGAACAGATGGGTATTCGGTAGAGTTTCAGCCACAC 18404  
Qy 204 GluPheLeuProSerGlnHisGluLeuAsnPheValProAlaAspProLeuThrMetAla 223  
Db 18403 GAGCCCGCTCTGGACACGACGAAATGACCTTCGTACGCCGATCGCTGACGATGCC 18344  
Qy 224 AspArgHisIleAlaLysHisGlyValArgGluMetAlaGluGlnSerGlyMetVal 243  
Db 18343 GACAACATCATGACCTTCGGGTAGTATTCGGGAGATCGTTCCTCAAGGGATCAAG 18284  
Qy 244 AlaThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHisMet 263  
Db 18283 GCCACCTTCATCGCTTAAGCCTTTTACCGATCACCGCGGTGAGTATGCTTCACGTC 18224  
Qy 264 SerLeuGlnAspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAspGlyTyrGly 283  
Db 18223 TCCCTCTTTGAGGGGGACAC-----AATGCCCTTCTATGACGCGGTGACGAGCGCTCGC 18170  
Qy 284 MetSerThrLeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyrValProGluAlaThr 303  
Db 18169 ATGTCCGCGGTGGCAAGTATTCGTCCCGGTCTTTTGACCATGCTCCGAGATCATC 18110  
Qy 304 TyrPhePheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAlaPro 323  
Db 18109 GCCGTGACGAACAGATGGTTAACTCTACCGCTTTCCGCTGGGGAGAGCGCT 18050  
Qy 324 ThrLysCysCysTyrAlaIleAspAsnArgThrSerAlaPheArgLeu-----Cys 340  
Db 18049 AACTACATTTGCTGGGTGCGCAATAATCGTTGACACTGGTTCGTATTCGATGATAAG 17990  
Qy 341 AsnSerLysSerGluGlyIleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsnPro 360  
Db 17989 CCCGACAGGTTCTCGGTCGGTGAGATGTTGCTTATCGATTCTGTCGCCAACCCA 17930  
Qy 361 TyrLeuAlaPheSerAlaIleAlaAlaGlyIleSerGlyIleGluGluLysLeuGlu 380  
Db 17929 TATCTGGCGTATCTCTGGTACTTGTCTGTTTGGACGCGCATCGAGAGGAGCTACCG 17870  
Qy 381 LeuProProAlaSerGlyAsnValTyr-----AsnAspLysGlu----- 394  
Db 17869 CTTCCAGAGGAGGATCCGACGACGTGTGGCAGTTGTTCAGCTCCGCGAGCTGACGCACTG 17810  
Qy 395 ---LeuProGluPheProAsnSerLeuGlnAsnAlaThrHisLeuLeuLysGluSerLys 413  
Db 17809 GGGATCAAGCGCTGCCGAAAGTCTGGGTGCTCGGATCCGATCCGATGGAGGAGCTCGAG 17750  
Qy 414 MetLeuAsnLysThrPheGlyGluLysLeuIleLeuHisTyrValAsnAlaAlaAsnVal 433

Db 17749 CTGCTGCTGAGACCTCTCGCGAACACGCTACGATTATTCTTACGCAATAAGAGGCG 17690  
Qy 434 GluIleAsnGluPheSerLysGlnValThrAspTrpGluLeuAsnGln 449  
Db 17689 GAATTCAGGAGTACAAACCGCAGGTTCAGCAGTTGCGCTTATCGC 17642

## RESULT 6

ABQ82091  
ID ABQ82091 standard; DNA; 5500 BP.

XX ABQ82091;

XX 22-NOV-2002 (first entry)

DE Brevibacterium lactofermentum glnA2 and glnE encoding DNA SEQ ID NO:1.

XX Brevibacterium lactofermentum; glnA2; glnE; L-glutamine; fermentation;  
KW coryneform bacterium; glutamine synthetase adenyl transferase;  
KW glutamine synthetase; liver function promoting agent; enzyme; seasoning;  
KW gene; ds.

XX Brevibacterium lactofermentum.

XX Key Location/Qualifiers  
CDS 659..1999  
FT CDS /\*tag= a  
FT /\*product= "glnA2"  
FT 2066..5203  
FT /\*tag= b  
FT /\*product= "glnE"

XX EPI1229121-A2.

XX 07-AUG-2002.

XX 05-FEB-2002; 2002EP-0001993.

XX 05-FEB-2001; 2001JP-0028163.

PR 30-MAY-2001; 2001JP-0162806.

XX (AJIN ) AJINOMOTO CO INC.

XX Nakamura J, Izui H, Moriguchi K, Kawashima H, Nakamatsu T;  
PI Kurahashi O;

XX WPI; 2002-629685/68.

DR P-PSDB; ABP53500, ABP53501.

XX Coryneform bacterium which has L-glutamine producing ability and has  
PT been modified so that its intracellular glutamine synthetase activity  
PT should be enhanced, useful for producing L-glutamine

XX Claim 11; Page 19-26; 39pp; English.

XX The present invention describes a coryneform bacterium (I) which has  
CC L-glutamine producing ability and has been modified so that its  
CC intracellular glutamine synthetase activity should be enhanced. Also  
CC described is a DNA (II) coding for a protein having glutamine synthetase  
CC activity or glutamine synthetase adenyl transferase activity (see  
CC ABP53500 and ABP53501 respectively). (I) is useful for producing  
CC L-glutamine, by culturing a bacterium in a medium to produce and  
CC accumulate L-glutamine in the medium and collecting the L-glutamine.  
CC L-glutamine produced by (I) is useful industrially as an ingredient of  
CC seasonings, as liver function promoting agents, in amino acid  
CC transfections, and in comprehensive amino acid preparation. (II) is  
CC useful for breeding (I). The by-product of L-glutamic acid is  
CC suppressed and the production efficiency of L-glutamine is improved  
CC using (II). The present sequence encodes Brevibacterium lactofermentum  
CC glnA2 and glnE, which have glutamine synthetase activity and glutamine  
CC synthetase adenyl transferase activity respectively.

XX Sequence 5500 BP; 1241 A; 1291 C; 1590 G; 1378 T; 0 other;



XX The present invention provides a number of nucleotide and protein  
 CC sequences from the Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of coryneform bacterium, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from coryneform bacterium, and identifying a homologue of a gene derived  
 CC from coryneform bacterium. Corynebacterium bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a nucleic acid described  
 CC in the exemplification of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.

XX Sequence 1338 BP; 336 A; 367 C; 343 G; 302 T; 0 other;

## Alignment Scores:

Pred. No.: 9,41e-41 Length: 1338  
 Score: 470.50 Matches: 134  
 Percent Similarity: 47.08% Conservative: 84  
 Best Local Similarity: 28.94% Mismatches: 174  
 Query Match: 19.80% Indels: 71  
 DB: 22 Gaps: 12

US-10-098-602A-2 (1-454) x AAH67407 (1-1338)

Qy 15 GlyLysIleAspThrValValLeuAlaCysValAspMetGlnGlyArgLeuMetGlyLys 34  
 Db 85 GCCACTTGAAGTCAGTGGTGTGGCTCTCCAGAACTAGAGTCGGTTG----- 135  
 Qy 35 ArgLeuThrGlyArgHisPheLeuGlyLeuAspGlnLysIleSerIleSerThrPhe 54  
 Db 136 -----GAAGAAGGCTCGGATTTCGAT----- 156  
 Qy 55 ValTyrAlaValThrIleGluGlyIleAlaGlyGlyTyrGluIleSerSerValAsp 74  
 Db 157 -----GGCTCAGCCATTGAGGGGTACGCG-----CGTATCTCGAAGCGGAC 198  
 Qy 75 ThrGlyTyrSerAspCysHisLeuCysAlaAspLeuAsnSerLeuHisLeuProTrp 94  
 Db 199 ACC-----ATTGCGCGCCAGATCATCGATTCAGGTCCTCCCACTA 243  
 Qy 95 SerGluGly-----AlaValLeuAlaIleSerAsnProHisAsnPheValThr----- 110  
 Db 244 GAAGCGGCATCTCAAACTCAGCAGCAGCCGCTGTTTCGGATGTCACGATCCAGAC 303  
 Qy 111 SerGluProLeuPheCysSerProArgValIleLeuMetGlnGlnIleGluArgLeuAla 130  
 Db 304 GGACAGCCATCTTTTCTGACCCGCGCAAGTGTGCGCAGCAGGTCCAACTAGCTGCA 363  
 Qy 131 AsnLeuLysLeuLysGlyLeuPheAlaSerGluLeuGluPheAsnLeuPheAsnGluThr 150  
 Db 364 GATGAAGCTTGACCTGCATCATCTCACCAGAGATTGAGTTCTATTGTTG----- 414  
 Qy 151 TyrLysSerAlaSerGlnLysHisIleTrpLysAsnLeuLysThr----- 164  
 Db 415 -----CAAGCCTTCGACCAACGAGTCCACCTGTG 447  
 Qy 165 AlaGlnProHisHisGlnTrpMetAsnIleSerAlaSerSerGlyIleGluThrPheMet 184  
 Db 448 CCCACTGACACCGCGGATATTTCGACCAAGCCCATTCATGAGGCGCGCAATTCCTGT 507  
 Qy 185 ArgSerValArgAsnLysLeuGluGluAlaGlyIleLeuMetGluAlaThrHisProGlu 204  
 Db 508 CGAAACCGCATGGTAGCGCTCGAGGAATCTCGCATCCCTGTCGAGTTCTCCCAACATGAA 567  
 Qy 205 PheLeuProSerGlnHisGluLeuAsnPheValProAlaAspProLeuThrMetAlaAsp 224  
 Db 568 ACTGCATGCGCCAGCAAGAATCGATTAGCCATGCGGATGCGCTCACCATGCGCCAC 627  
 Qy 225 ArgHisIleLeuAlaLysHisGlyValArgGluMetAlaGluGlnSerGlyMetValAla 244

Db 628 AACATCATGACCTTCGCTACATCATGAAACAGGTGGCAAGGGACCAAGCGCTCGGGGCA 687  
 Qy 245 ThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSer 264  
 Db 688 TCATTATGCCAAGCCATTCCAAAGACATGCAGGCTCCGCCATGCACCGCACATGTCC 747  
 Qy 265 LeuGlnAspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMet 284  
 Db 748 TTATTTGAGGGCGATACC-----ACGCGTTCACGATCCAGACGATCTTACATGCTG 801  
 Qy 285 SerThrLeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyrValProGluAlaThrTyr 304  
 Db 802 TCCAAACCGCAAAACAGATTCTACGCTGGAAATCTTCATCAGCTCAGAAATTCACCGCT 861  
 Qy 305 PhePheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThr 324  
 Db 862 GTGACCAACAGTGGGTCAATTCCTACAAAGCATCGTGTACGAAACGAGCTCCAACT 921  
 Qy 325 LysCysCysTrpAlaIleAspAsnArgThrSerAlaPheArgLeuCysAsn----- 341  
 Db 922 GCGCAACCTGGGGTGTATCTAATCGTTCGCGCTGGTTCGTTCCTACCTACCGTTTG 981  
 Qy 342 SerLysSerGluGlyIleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsnProTyr 361  
 Db 982 AATAAGGAGGAGTGGCGCGGTGGAGTGGCTTCTCTCATACCGCTTGTAAACCCATAT 1041  
 Qy 362 LeuAlaPheSerAlaIleAlaAlaGlyIleSerGlyIleGluGlyLysLeuGluLeu 381  
 Db 1042 TTGGCGTTTTCAGTGATGCTCGGCGCTGGTTTGAAGGCATTAAGAAGGTTATGAGCTC 1101  
 Qy 382 ProProProlaSerGlyAsnVal----- 389  
 Db 1102 GACGAGCAGCTGAGGACGATATCTCCAACTTCGAGCTTCGGAACGTCGCGCCATGGC 1161  
 Qy 390 TyrAsnAspLysGluLeuProGluPheProAsnSerLeuGlnAsnAlaThrHisLeuLeu 409  
 Db 1162 TACAACGAT-----CTGCCAAGCAGCCTTGCATCAGGACATCGCGCAAAATG 1206  
 Qy 410 LysGluSerLysMetLeuAsnLysThrPheGlyGlyLysLeuIleLeuHisTyrValAsn 429  
 Db 1207 GAAAGTCAGAGCTGTGTGTCATCTCGTGAGCAGCTTTTGTAGTTTTTCTTGTGCG 1266  
 Qy 430 AlaAlaAsnValGluIleAsnGluPheSerLysGlnValThrAspTrpGluLeuAsnGln 449  
 Db 1367 AATAAGTGGCGTGAATGGCGTGACTCAAGAGCAGATCACTCGTGGGAGCTCCGAAAC 1326  
 Qy 450 GlyPheAsn 452  
 Db 1327 AATCTTGAT 1335  
 RESULT 8  
 AAF71800  
 ID AAF71800 standard; DNA; 1461 BP.  
 XX AAF71800;  
 AC AAF71800;  
 XX 30-APR-2001 (first entry)  
 DT  
 XX  
 DE Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:95.  
 KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;  
 KW fine chemical production; microorganism; organic acid; nucleoside;  
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;  
 KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;  
 KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.  
 OS Corynebacterium glutamicum.  
 XX  
 PN W0200100843-A2.  
 XX  
 PD 04-JAN-2001.  
 XX  
 PF 23-JUN-2000; 2000WO-IB00923.



Qy 265 LeuGlnAspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMet 284  
 Db 848 TTTATTTAGGGCGGATACC-----AACGGTTCCACGATCCAGACGATTTCTTACATGCTG 901  
 Qy 285 SerThrLeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyrValProGluAlaThrTyr 304  
 Db 902 TCCAAACCGCAACACAGTTTCATCGTGAATCTTGCATCAGCTCCAGAAATTCACCGCT 961  
 Qy 305 PhePheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThr 324  
 Db 962 GTGACCAACCAAGTGGTCAATTCCTACAAACGCATCGTGTACGGAAACGAGCTCCAACT 1021  
 Qy 325 LysCysCysTrpAlaIleAspAsnArgThrSerAlaPheArgLeuCysAsn----- 341  
 Db 1022 CGCGCAACCTTGGGTGTATCAATCGTTCTCGCGTGGTTCGTCTCCTACCTACCGTTG 1081  
 Qy 342 SerLysSerGluGlyIleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsnProThr 361  
 Db 1082 AATAAGAGAGAGTCCGCGGTGGAGTGGTCTTCTGTATACCGTTGTAAACCCATAT 1141  
 Qy 362 LeuAlaPheSerAlaIleAlaAlaGlyIleSerGlyIleGluGlyLeuLysLeuGluLeu 381  
 Db 1142 TTGCGTTTTCAGTGATCTCGCGCTCGGTGTTTGAAGGCATTAAGAGGTTATGAGCTC 1201  
 Qy 382 ProProAlaSerGlyAsnVal----- 389  
 Db 1202 CACGAGCCAGCTGAGGAGCATATCTCAACTTGAGCTTCCGGAAACGTCGCGCATGGC 1261  
 Qy 390 TyrAsnAspLysGluLeuProGluPheProAsnSerLeuGlnAsnAlaThrHisLeuLeu 409  
 Db 1262 TACAACCAT-----CTGCCAAGCAGCCTTGTATCAGGCACTGCGCAAAATG 1306  
 Qy 410 LysGluSerLysMetLeuAsnLysThrPheGlyGlyLysLeuIleLeuHisTyrValAsn 429  
 Db 1307 GAAGAATCAGAGCTGTGTGTGACATCTCGTGAGCAGCATTTTGTAGTTTCTTGGCG 1366  
 Qy 430 AlaAlaAsnValGluLeuAsnGluPheSerLysGlnValThrAspTrpGluLeuAsnGln 449  
 Db 1367 AATAAGTGGCGTGAATGGCGTGACTACCAAGAGCAGATCACTCGTGGGAGCTCCGAAAC 1426  
 Qy 450 GlyPheAsn 452  
 Db 1427 AATCTTGAT 1435

RESULT 9  
 AAH68531/c  
 ID AAH68531 standard; DNA; 349980 BP.  
 XX  
 AC AAH68531;  
 XX  
 DT 26-SEP-2001 (first entry)  
 XX  
 DE C glutamicum coding sequence fragment SEQ ID NO: 7066.  
 XX  
 KW Corynebacterium; amino acid synthesis; vitamin; saccharide;  
 KW organic acid synthesis; ds.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN EF1108790-A2.  
 XX  
 PD 20-JUN-2001.  
 XX  
 PF 18-DEC-2000; 2000EP-0127688.  
 XX  
 PR 16-DEC-1999; 99JP-0377484.  
 PR 07-APR-2000; 2000JP-0159162.  
 PR 03-AUG-2000; 2000JP-0280988.  
 XX  
 PA (KYOW ) KYOWA HAKKO KOGYO KK.  
 XX  
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI; 2001-376931/40.  
 XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene  
 XX  
 XX Disclosure; SEQ ID NO: 7066; 246pp + Sequence Listing; English.  
 XX  
 XX The present invention provides a number of nucleotide and protein  
 CC sequences from the Corynebacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of corynebacterium, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from Corynebacterium, and identifying a homologue of a gene derived  
 CC from corynebacterium. Corynebacterium bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a nucleic acid described  
 CC in the exemplification of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.  
 XX  
 SQ Sequence 349980 BP; 80724 A; 98367 C; 90490 G; 80399 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 4,55e-37 Length: 349980  
 Score: 470.50 Matches: 134  
 Percent Similarity: 47.08% Conservative: 84  
 Best Local Similarity: 28.94% Mismatches: 174  
 Query Match: 19.80% Indels: 71  
 DB: 22 Gaps: 12  
 US-10-098-602A-2 (1-454) x AAH68531 (1-349980)  
 Qy 15 GlyLysIleAspThrValValLeuAlaCysValAspMetGlnGlyArgLeuMetGlyLys 34  
 Db 264071 GGCACCTTGAAGTCAGTGGTGTGGCTCTCGCAGAACTAGAGTCGCGTTG----- 264021  
 Qy 35 ArgLeuThrGlyArgHisPheLeuGlyLeuAspGlnLysLysIleSerIleSerThrPhe 54  
 Db 264020 -----GAAGAAGCATCGGATTCGAT----- 264000  
 Qy 55 ValTyrAlaValThrIleGluGlyIleAlaGlyGlyTyrGluIleSerSerValAsp 74  
 Db 263999 -----GGCTCAGCCATTGAGGCTACGCG-----CGTATCTCGAAGCCGAC 263958  
 Qy 75 ThrGlyTyrSerAspCysHisLeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTrp 94  
 Db 263957 ACC-----ATTGCCCGCCAGATCCATCCAGATTCAGGTCCTCCACTA 263913  
 Qy 95 SerGluGly---AlaValLeuAlaIleSerAsnProHisAsnPheValThr----- 110  
 Db 263912 GAAGCGGGCATCTCAAAACTGACGAGCAGCGCTGTGTTTCGATGTCACGATGCCAGAC 263853  
 Qy 111 SerGluProLeuPheCysSerProArgValIleLeuMetGlnGlnIleGluArgLeuAla 130  
 Db 263852 GGACGCATCTTTTCTGACCCCGCCCAAGTCTGCGCAGGAGGTCCTCACTAGCTGCA 263793  
 Qy 131 AsnLeuLysLeuLysGlyLeuPheAlaSerGluLeuGluPheAsnLeuPheAsnGluThr 150  
 Db 263792 GATGAAGGCTTGACCTGCATGATCTCACCAGAGATTGAGTTCTATTGGTG----- 263742  
 Qy 151 TyrLysSerAlaSerGlnLysHisTrpLysAsnLeuLysThr----- 164  
 Db 263741 -----CAAAGCCTTGCACCAACGACGACGACCTCCACCTGTG 263709  
 Qy 165 AlaGlnProHisGlnTrpMetAsnIleSerAlaSerSerGlyIleGluThrPheMet 184  
 Db 263708 CCCACTGACACCGCGGATATTTCCACCAAGCCACATTTCAATGAGGCGCGCAATTTCCGT 263649  
 Qy 185 ArgSerValArgAsnLysLeuGluAlaGlyIleLeuMetGluAlaThrHisProGlu 204









CC used in gene therapy. Antibodies to (I) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins.

XX Sequence 1344 BP; 426 A; 224 C; 277 G; 417 T; 0 other;

## Alignment Scores:

Pred. No.: 5-51e-38 Length: 1344  
 Score: 445.00 Matches: 131  
 Percent Similarity: 44.14% Conservative: 80  
 Best local Similarity: 27.41% Mismatches: 213  
 Query Match: 18.73% Indels: 54  
 DB: 24 Gaps: 10

US-10-098-602A-2 (1-454) x ABN68493 (1-1344)

QY 1 MetThrIleThrTyrAspGluLeuAsnAsnLeuIleArgAsnGlyLysIleAspThrVal 20  
 DB 1 ATGACTATCACAGCAGAGATATTCGCGAAGTTAAGGAAAAAATGTTACCTTCCTC 60  
 QY 21 ValLeuAlaCysValAspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHis 40  
 DB 61 CGTTTGATGTTACAGATATCTTAGGAGTAATGAAAAATGTCGAAATCCG----- 111  
 QY 41 PheLeuGlyLeuAspGln-----LysLysIleSerIleSerThrPheValTyrAla 57  
 DB 112 -----GCTACAGATGCAACAAATAGACAAAGTTTATCAATAAAGCAATGTTGATGGA 165  
 QY 58 ValThrIleGluGlyIleAlaGlyGlyGlyTyrGluIleSerSerValAspThrGlyTyr 77  
 DB 166 TCTTCTATTGAGGCGCTT-----GTTGCTATCATGAA 198  
 QY 78 SerAspCysHisLeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTrpSerGlu--- 96  
 DB 199 TCGATATGATTTATATCTGATTTAGTACTGGAATGCTTCTTGGGAGATGAA 258  
 QY 97 ---GlyAlaValLeuAlaIleSerAsnProHisAsnPheValThrSerGluProLeuPhe 115  
 DB 259 AATGGAGCTGTTGCGAGTCTAATCTGTGATATCTATACAGCAGAGAGAACTTTTGCC 318  
 QY 116 CysSerProArgValIleLeuMetGlnGlnIleGluArgLeuAlaAsnLeuLysLeuLys 135  
 DB 319 GGAATCCAGCGTGGAAATCTTAAGCGTAATATGAAACGTAATGCAAGAGATGGGCTATAAG 378  
 QY 136 GlyLeu---PheAlaSerGluLeuGluPheAsnLeuPheAsnGluThrTyrLysSerAla 154  
 DB 379 TCATTTAACTTAGGACAGCACTGAAATTTTCTTATTT----- 417  
 QY 155 SerGlnLysHisTrpLysAsnLeuLysThrAlaGlnProHisGlnTrpMetAsnIle 174  
 DB 418 -----AAAATGGATGAAATGGTAACCAACCAACACTTGATGTT 453  
 QY 175 SerAlaSerSerGly-----IleGluThrPheMetArg 185  
 DB 454 AATGACAAAAGGAGGCTACTTTGACTTAGCTCCACAGACTTAGCAGATATAATACAGTCGT 513  
 QY 186 SerValArgAsnLysLeuGluAlaGlyIleLeuMetGluAlaThrHisProGluPhe 205  
 DB 514 GAAATTTGTAATGCTTTACACAAATGGGCTTGAAGTTGAAGCTAGTCATCATGAAGTA 573  
 QY 206 LeuProSerGlnHisGluLeuAsnPheValProAlaAspProLeuThrMetAlaAspArg 225  
 DB 574 GCAGTTGGACAGCATGAAATGACTTTAAATATGATGATGTTTAAAGCCTGTGACAAC 633  
 QY 226 HisIleIleAlaLysHisGlyValArgGluMetAlaGluSerGlyMetValIleThr 245  
 DB 634 ATTCAATTATTTAAATTTGGTTGTTAAACAATTCGCTGAAGCATGGTTTATATGCTACT 693  
 QY 246 PheMetAlaLysLeuSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeu 265  
 DB 694 TTTATGGCAAAACCTAAGTTTGGTATCAATGGTTGAGGTATGCAATGTAATATGCTCT 753  
 QY 266 GlnAspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMet--- 284

DB 754 TTTGATAATGAA---GGGAACAATCTTTCTTTGATCCAGAAGATCCACGTCGAATGCAA 810  
 QY 285 ---SerThrLeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyrValProGluAlaThr 303  
 DB 811 TTATCAGAGGATGCTACTCTTTTAGTGGTTTAATGAACACGCTTACAAATTACACA 870  
 QY 304 TyrPhePheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAlaPro 323  
 DB 871 GCTATCATTAATCAACCTGTTAATCTTTATAACGTTTGGTACCAGGATACGAAGCACA 930  
 QY 324 ThrLysCysCysTrpAlaIleAspAsnArgThrSerAlaPheArgLeuCysAsnSerLys 343  
 DB 931 GTTTACGTTGCTTGGCAGGTCGCAATCGTTCCGCCACTTATTCTGTTGCTGCTCATCGA 990  
 QY 344 SerGluGlyIleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsnProTyrLeuAla 363  
 DB 991 GGTATGGGAACACGCTAGNAATTCGCTCGGTTGACCCACAGCAAAATCCTTATCTTGT 1050  
 QY 364 PheSerAlaIleAlaAlaGlyIleSerGlyIleGluGlyLysLeuGluLeuProPro 383  
 DB 1051 TTATCAGTACTTTGGGATCAGGACTTCAGGGAATCGAATAAATTTGACACAGNA 1110  
 QY 384 ProAlaSerGlyAsnValTyr-----AsnAspLysGluLeuPro 396  
 DB 1111 CCAATTGAGACAAATATCTATGCTATGACTGTAGAAGAACGTCGTCACAGCAGGTATTGTT 1170  
 QY 397 GluPheProAsnSerLeuGlnAsnAlaThrHisLeuLeuLysGluSerLysMetLeuAsn 416  
 DB 1171 GATTTACCATCTACTCTTCATAATGCTCTTGAAGCAGCTTGAAGAGATGAAGTGGTAAAA 1230  
 QY 417 LysThrPheGlyGluLysLeuIleLeuHisTyrValAsnAlaAlaAsnValGluLeuAsn 436  
 DB 1231 GCGGCCCTAGGAACCTCATATCTATACTAATTTCTTAGACGCTAAGCGTATTGAATGGCA 1290  
 QY 437 GluPheSerLysGlnValThrAspTrpGluLeuAsnGlnGlyPheAsnArgTyr 454  
 DB 1291 AGTTATGCGACTTATGTTTTCACATGGGAAATTTGATAATTACCTAGATTATAT 1344

RESULT 14  
 ABN71527/c  
 ID ABN71527 standard; DNA; 2155561 BP.  
 XX AC ABN71527;  
 XX DT 02-JUL-2002 (first entry)  
 XX DE Streptococcus polynucleotide SEQ ID NO 10967.  
 XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 XX KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;  
 XX KW antinflammatory; infection; vaccine; meningitis; gene therapy; ds.  
 XX OS Streptococcus sp.  
 XX PN WO200234771-A2.  
 XX PD 02-MAY-2002.  
 XX PF 29-OCT-2001; 2001WO-GB04789.  
 XX PR 27-OCT-2000; 2000GB-0026333.  
 XX PR 24-NOV-2000; 2000GB-0028727.  
 XX PR 07-MAR-2001; 2001GB-0005640.  
 XX PA (CHIR-) CHIRON SPA.  
 XX PA (GENO-) INST GENOMIC RES.  
 XX PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;  
 XX PI Tettelin H;  
 XX DR WPI; 2002-352536/38.

PT New Streptococcus protein for the treatment or prevention of infection  
 PT or disease caused by Streptococcus bacteria, such as meningitis, and  
 PT for detecting a compound that binds to the protein -

PS Claim 8; Page 4196-4488; 4525pp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in  
 CC the specification. The proteins have antibacterial and antiinflammatory  
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
 CC antibodies that bind (I) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
 CC biological sample. (I) is used to determine whether a compound binds to  
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (I) may be used to recombinantly produce (I) and may be  
 CC used in gene therapy. Antibodies to (I) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins.

XX SQ Sequence 2155561 BP; 695741 A; 385320 C; 382958 G; 691542 T; 0 other;

#### Alignment Scores:

Pred. No.: 4.23e-33 Length: 2155561  
 Score: 445.00 Matches: 131  
 Percent Similarity: 44.14% Conservative: 80  
 Best Local Similarity: 27.41% Mismatches: 213  
 Query Match: 18.73% Indels: 54  
 DB: 24 Gaps: 10

US-10-098-602A-2 (1-454) x ABN71527 (1-2155561)

Qy 1 MetThrIleThrTyAspGluLeuAsnLeuIleArgAsnGlyLysIleAspThrVal 20  
 Db 1726056 ATGACTATACAGCAGAGATATCGTCGCGAAGTAAAGAAAAATGTTACCTTCCTC 1725997  
 Qy 21 ValLeuAlaCysValAspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHis 40  
 Db 1725996 CGTTTGATGTTTCACAGATATCTTAGGAGTAAAGAAAAATGTCGAATTCGCG- 1725946  
 Qy 41 PheLeuGlyLeuAspGln-----LysLysIleSerIleSerThrPheValTyrAla 57  
 Db 1725945 -----GCTACAGATGACAAATTAGACAAAGTTTATCAATAAGCAATGTTGATGGA 1725892  
 Qy 58 ValThrIleGluGlyIleAlaGlyGlyGlyTyrGluIleSerSerValAspThrGlyTyr 77  
 Db 1725891 TCTTCTATTGAGGCTTT-----GTTGCTATCAATGAA 1725859  
 Qy 78 SerAspCysHisIleuCysAlaAspLeuAsnSerLeuHisIleuLeuProTrpSerGlu--- 96  
 Db 1725858 TCTGATATGATTTATATATCTTCGATTTAGATATCTTGGATGTTGCTTCTCGGAGATGAA 1725799  
 Qy 97 ---GlyAlaValLeuAlaIleSerAsnProHisAsnPheValThrSerGluProLeuPhe 115  
 Db 1725798 AATGGAGCTGTTGCGAGTCTATCTGTGATATCTATACAGCAAGAGAGAACCTTTTGCC 1725739  
 Qy 116 CysSerProArgValIleLeuMetGlnGlnIleGluArgLeuAlaAsnLeuLysLys 135  
 Db 1725738 GGAGATCCACGTCGAAATCTTAAAGCGTAATATGAACGATGCAAGAGATGGGCTATAAG 1725679  
 Qy 136 GlyLeu---PheAlaSerGluLeuGluPheAsnLeuPheAsnGluThrTyrLysSerAla 154  
 Db 1725678 TCATTTAACTTAGACCAACGACCTGGAATTTTCTCTATT----- 1725640  
 Qy 155 SerGlnLysHisTyrPheAsnLeuLysThrAlaGlnProHisHisGlnTrpMetAsnIle 174  
 Db 1725639 -----AAATGGATGAAATGTAACCAACACTTGATGTT 1725604  
 Qy 175 SerAlaSerSerGly-----IleGluThrPheMetArg 185

Db 1725603 AATGACAAAGGAGGCTACTTTTGACTTAGCTCCAAACAGACTTAGCAGATAATACACGTCGT 1725544  
 Qy 186 SerValArgAsnLysLeuGluAlaGlyLeuMetGluAlaThrHisProGluPhe 205  
 Db 1725543 GAAATTTGTTATGTCCTTAACAAATGGCTTTGAAGCTAGCTATCATCATGAAGTA 1725484  
 Qy 206 LeuProSerGlnHisGluLeuAsnPheValProAlaAspProLeuThrMetAlaAspArg 225  
 Db 1725483 GCAGTTGGACAGCATGAATTTGACTTTAAATATGATGATGATGTTTAAAGCCTGTGACAA 1725424  
 Qy 226 HisIleIleAlaLysHisGlyValArgGluMetAlaGlnSerGlyMetValAlaThr 245  
 Db 1725423 ATTCAAATTTATTAATTTGTTTAAACAATTCCTCGTAGCATGCTTTATATGCTACT 1725364  
 Qy 246 PheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeu 265  
 Db 1725363 TTTATGCAAAACCTTAAGTTTGGTATCAATGGTTTCAGGTATGCTATGTAATATATCTCTT 1725304  
 Qy 266 GlnAspAlaGluThrGluLysAsnAlaPheTyAspGlnAsnAspGluTyrGlyMet--- 284  
 Db 1725303 TTTGATAATGAA---GGGAACAATGCTTTCTTTGATCCAGAGATCCACGTTGGAATGCA 1725247  
 Qy 285 ---SerThrLeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyrValProGluAlaThr 303  
 Db 1725246 TTTATCAGAGATGCTACTACTTTTATAGTGGGTTAATGAACACGCTTACAATTACACA 1725187  
 Qy 304 TyrPhePheAlaSerTyrIleAsnSerTyrLysArgGluGlnProLeuThrPheAlaPro 323  
 Db 1725186 GCTATCATTTAATCCAACTGTTAATTTCTTATAAACGTTTGTGTACAGATACGAAGACCA 1725127  
 Qy 324 ThrLysCysCysTrpAlaIleAspAsnArgThrSerAlaPheArgLeuCysAsnSerLys 343  
 Db 1725126 GTTTACGTTGCTTGGGCGAGTCCCAATCGTTGCCACTTATTCGTGTACCTGATCTCGA 1725067  
 Qy 344 SerGluGlyIleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsnProTyrLeuAla 363  
 Db 1725066 GGTATGGAAACACGCTAGAAATTCGCTCGCTGACCCAAACAGCAAAATCCTTATCTTCT 1725007  
 Qy 364 PheSerAlaIleIleAlaGlyIleSerGlyIleGluGluLysLeuGluLeuProPro 383  
 Db 1725006 TTTATGATGTTTGGATCAGGATGAGGGAATCGAGAAATAAATTTGAAGCACCAAGAA 1724947  
 Qy 384 ProAlaSerGlyAsnValTyr-----AsnAspLysGluLeuPro 396  
 Db 1724946 CCAATTGACAAATATCTATGCTATGCTGTAGAGAAGCTGCTCAGCAGGTATTTGTT 1724887  
 Qy 397 GluPheProAsnSerLeuGlnAsnAlaThrHisLeuLeuLysGluSerLysMetLeuAsn 416  
 Db 1724886 GATTTACCATCTACTCTTCAATAATGCTTCTTGAAGCACTTGAAGAAGATGAAGTGTAAAA 1724827  
 Qy 417 LysThrPheGlyGluLysLeuIleLeuHisTyrValAsnAlaAlaAsnValGluIleAsn 436  
 Db 1724826 GCGGCCCTAGGAACCTCATATCTATATAATTTCTTAGACGCTAGGCGTATTTGAATGGCA 1724767  
 Qy 437 GluPheSerLysGlnValThrAspTyrGluLeuAsnGlnGlyPheAsnArgTyr 454  
 Db 1724766 AGTTATGCGACTTATGTTTCACAATGGGAATTTGATAATTACCTAGATTATAT 1724713

#### RESULT 15

ABN68494  
 ID ABN68494 standard; DNA; 1344 BP.

XX AC ABN68494;

XX 01-JUL-2002 (first entry)

XX Streptococcus polynucleotide SEQ ID NO 4901.

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 XX group A streptococcus; Streptococcus pyogenes; antibacterial; gene;  
 XX antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.

OS Streptococcus pyogenes.

XX WO200234771-A2.

XX 02-MAY-2002.

XX 29-OCT-2001; 2001WO-GB04789.

XX 27-OCT-2000; 2000GB-0026333.

XX 24-NOV-2000; 2000GB-0028727.

XX 07-MAR-2001; 2001GB-0005640.

XX (CHIR-) CHIRON SPA.

XX (GENO-) INST GENOMIC RES.

XX Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;

XX Tettelin H;

XX WPI; 2002-352536/38.

XX P-PSDB; ABP27863.

XX New Streptococcus protein for the treatment or prevention of infection

XX or disease caused by Streptococcus bacteria, such as meningitis, and

XX for detecting a compound that binds to the protein -

XX Claim 7; Page 3652; 4525pp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B

XX streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS

XX (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in

XX the specification. The proteins have antibacterial and anti-inflammatory

XX activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and

XX antibodies that bind (II) are used in the manufacture of medicaments for

XX the treatment or prevention of infection or disease caused by

XX Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.

XX Nucleic acids encoding (I) are used to detect Streptococcus in a

XX biological sample. (II) is used to determine whether a compound binds to

XX (I). A composition comprising (I) or a nucleic acid encoding (I), may be

XX used as a vaccine or diagnostic composition. The disease caused by

XX Streptococcus that is prevented or treated may be meningitis. Nucleic

XX acid encoding (II) may be used to recombinantly produce (I) and may be

XX used in gene therapy. Antibodies to (I) are used for affinity

XX chromatography, immunoassays, and distinguishing/identifying

XX Streptococcus proteins.

XX Sequence 1344 BP; 419 A; 219 C; 288 G; 418 T; 0 other;

SQ

Alignment Scores:

Pred. No.: 1.5e-37 Length: 1344  
Score: 441.00 Matches: 131  
Percent Similarity: 44.92% Conservative: 81

Best Local Similarity: 27.75% Mismatches: 218  
Query Match: 18.56% Indels: 42  
DB: 24 Gaps: 10

US-10-098-602a-2 (1-454) x ABN68494 (1-1344)

QY 1 MetThrIleThrTyrAspGluLeuAsnAsnLeuIleAargAsnGlyIysIleAspThrVal 20

DB 1 ATGGCAATACAGTAGTCGATTCGTTGAAGTCAAGAAAAAATGTAACGTTTCTT 60

QY 21 ValLeuAlaCysValAspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHis 40

DB 61 CGCTTGATGTTCTACTGATATCATGCGGCTTATGAAAAATGGAGATTCCTGCACTAAA 120

QY 41 PheLeuGlyLeuAspGlnIysLysIleSerIleSerThrPheValTyrAlaValThrIle 60

DB 121 --GAACAGTTAGAC--AAAGTATTGCTCAACAGGTTATGTTGATGTTTCATCTATC 174

QY 61 GluGlyIleAlaGlyGlyTyrGluIleSerSerValAspThrGlyTyrSerAspCys 80

DB 175 GAAGGTTT-----GTACGGATCAATGAGTCAGATATG 207

QY 81 HisLeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTyrSerGlu-----GlyAla 98  
DB 208 TACCTTTACCCGATTTAGACACTTGGATTGTTTTTCCCTGGGAGTCAAAATGAGCA 267  
QY 99 ValLeuAlaIleSerAsnProHisAsnPheValThrSerGluProLeuPheCysSerPro 118  
DB 268 GTTGCAGGTTTAAATTTGTGATATTATACAGCAGAAAGAACCTTTTGCAGGAGATCCT 327  
QY 119 ArgValIleLeuMetGlnGlnIleGluArgLeuAlaAsnLeuLeuLysGlyLeu--- 137  
DB 338 AGAGAAATTTAAAGAGCCCTGAACACATGAACGAGATCGGCTACAAATCATTTAAT 387  
QY 138 PheAlaSerGluLeuGluPheAsnLeuPhe-----AsnGluThrTyr 151  
DB 398 CTTGACCCAGAACCAAGAAATTTTCCTTTTAAAGATGGATGATAAAGGTAATCCGACACT 447  
QY 152 LysSerAlaSerGlnLysHisTrpLysAsnLeuLysThrAlaGlnProHisHisGlnTrp 171  
DB 448 GAA-----GTTAACGATAAATGGTGTAT 471  
QY 172 MetAsnIleSerAlaSerSerGlyIleGluThrPheMetArgSerValArgAsnLysLeu 191  
DB 472 TTTGATTTTAGCCCAATTGACTTAGCAGACAAACACGCCCGGAAATTTGTGAATATTTTA 531  
QY 192 GluGluAlaGlyIleLeuMetGluAlaThrHisProGluPheLeuProSerGlnHisGlu 211  
DB 532 ACGAAATGGGTTTTGAAGTGAAGCTAGTCATCATGAAGTGGCTGTGGTCAACATGAG 591  
QY 212 LeuAsnPheValProAlaAspProLeuThrMetAlaAspArgHisIleIleAlaLysHis 231  
DB 592 ATTGATTTTAAATATGTCAGATGTTTGAAGCTGTGTGATAATATTCAAAATTTTAAAGCTA 651  
QY 232 GlyValArgGluMetAlaGlnSerGlyMetValAlaThrPheMetAlaLysLeuSer 251  
DB 652 GTTGTAAACAGGATTCGCCGTAACATGACATTTATGTACTTTTCATGGCTAAACAAA 711  
QY 252 SerThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeuGlnAspAlaGluThrGlu 271  
DB 712 TTTGGAATAGCTGGATGACATGTAACATGTCTTTGTTGATAACCAA---GGT 768  
QY 272 LysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMet-----SerThrLeuAlaArg 289  
DB 769 AATAATGCTTTTATGATGAGCTGATGAAGCGAGGATGCAGTTATCAGAAGATGCTTAT 828  
QY 290 AsnTrpIleAlaGlyLeuLeuLysTyrValProGluAlaThrTyrPhePheAlaSerTyr 309  
DB 829 TATTTCTTGGGAGGACTAATGAAGCATGCTTAACTACTACATGCTATCACTAACCCCTACA 888  
QY 310 IleAsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThrLysCysCysTrpAla 329  
DB 889 GTGAATCTTATAAACGATTAGTCCAGTTATGAGGCACCTGTTTATGTCGCTTGGGCT 948  
QY 330 IleAspAsnArgThrSerAlaPheArgLeuCysAsnSerLysSerGluGlyIleAsnVal 349  
DB 949 GGAAGTAATCGTTCACCGCTTATCGGTGTTCCAGCATCACGTGTTGGAACCGGCTTG 1008  
QY 350 GluLeuArgIleGlyGlyAlaAspLeuAsnProTyrLeuAlaPheSerAlaIleAla 369  
DB 1009 GAGTTACGTTCCGGTTAGTCCGACACATATCTTTATTTAGCCTTGGCTGCTTCTCTGGA 1068  
QY 370 AlaGlyIleSerGlyIleGluGlyLysLeuGluLeuProProProAlaSerGlyAsnVal 389  
DB 1069 GCTGGATTAGTGGTATCATTAACAAAATTTGAAGCTCCAGAACCCGCTTGAAGCTAACAT 1128  
QY 390 Tyr-----AsnAspLysGluLeuProGluPheProAsnSerLeu 402  
DB 1129 TATACCATGACAATGAAGAACGAAATGAAGCAGGCAATTTATGATTGCTTCATCAACGCTT 1188  
QY 403 GlnAsnAlaThrHisLeuLeuLysGluSerLysMetLeuAsnLysThrPheGlyGluLys 422  
DB 1189 CATATATGCTTAAAGCTCTTCAAAAAGATGATGTGTCACAAAAGGCACATAGTACCAT 1248  
QY 423 LeuIleLeuHisTyrValAsnAlaAlaAsnValGluIleAsnGluPheSerLysGlnVal 442

```

Db      1249  ATCTACACTAATTTCTTAGAAGCAAAACGAATTGAATGGTCTCTCTATGCAACTTTGTT 1308
Qv      443   ThrAspTrpGluLeuAsnGlnGlyPheAsnArgTyr 454
Db      1309  TCTCAATGGGAATTGACCATTTATTCATTAATTAT 1344

```

Search completed: December 18, 2003, 00:52:31  
Job time : 7163 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 22:44:55 ; Search time 31 Seconds  
(without alignments)  
2735.110 Million cell updates/sec

Title: US-10-098-602A-2  
Perfect score: 2376  
Sequence: 1 MTITYDELNNLRNGKIDTV.....INFSKQVTWELNQGPNRY 454

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 696363 seqs, 186758610 residues

Total number of hits satisfying chosen parameters: 696363

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	773.5	32.6	454	15	US-10-156-761-14253
2	475.5	20.0	446	15	US-10-062-458-2
3	470.5	19.8	446	10	US-09-738-626-5942
4	432	18.2	448	9	US-09-815-242-13378
5	427.5	18.0	453	15	US-10-156-761-13481
6	424.5	17.9	446	9	US-09-815-242-5733
7	424.5	17.9	446	9	US-09-815-242-12338
8	424.5	17.9	446	9	US-09-815-242-12735
9	402	16.9	451	9	US-09-815-242-10514
10	322	13.6	477	10	US-09-738-626-6985
11	291	12.2	469	15	US-10-156-761-13532
12	287	12.1	472	9	US-09-815-242-11137
13	281.5	11.8	500	15	US-10-187-267A-57
14	265.5	11.2	469	9	US-09-815-242-12088
15	265	11.2	481	9	US-09-815-242-11507

16	247	10.4	469	9	US-09-815-242-13789	Sequence 13789, A
17	241	10.1	468	12	US-10-299-799-1	Sequence 1, Appli
18	241	10.1	469	9	US-09-815-242-10418	Sequence 10418, A
19	238.5	10.0	306	15	US-10-156-761-7750	Sequence 7750, Ap
20	106	4.5	373	15	US-10-205-823-152	Sequence 152, App
21	96	4.0	318	12	US-10-306-762-211	Sequence 211, App
22	95.5	4.0	522	12	US-10-094-749-3166	Sequence 3166, Ap
23	94.5	4.0	728	9	US-09-815-242-5216	Sequence 5216, Ap
24	94	4.0	606	10	US-09-771-161A-240	Sequence 240, App
25	94	4.0	606	10	US-09-771-161A-241	Sequence 241, App
26	94	4.0	651	9	US-09-870-937-10	Sequence 10, Appl
27	94	4.0	651	10	US-09-974-298-112	Sequence 112, App
28	94	4.0	651	12	US-10-354-358-8	Sequence 8, Appli
29	93.5	3.9	6281	9	US-09-815-242-12996	Sequence 12996, A
30	93	3.9	1455	12	US-10-238-075-935	Sequence 935, App
31	92	3.9	390	11	US-09-214-592-28	Sequence 28, Appl
32	92	3.9	2086	9	US-09-815-242-5639	Sequence 5639, Ap
33	92	3.9	5795	9	US-09-815-242-12610	Sequence 12610, A
34	91	3.8	1050	11	US-09-954-987B-175	Sequence 175, App
35	91	3.8	1050	12	US-10-272-502A-22	Sequence 22, Appl
36	90	3.8	521	15	US-10-081-051-60	Sequence 60, Appl
37	89	3.7	2677	12	US-10-144-194A-22	Sequence 22, Appl
38	88	3.7	531	15	US-10-205-823-82	Sequence 82, Appl
39	87.5	3.7	441	12	US-10-396-122-79	Sequence 79, Appl
40	87.5	3.7	1185	9	US-09-815-242-11466	Sequence 11466, A
41	87	3.7	1294	9	US-09-815-242-13724	Sequence 13724, A
42	86.5	3.6	605	9	US-09-782-980-62	Sequence 62, Appl
43	86.5	3.6	1634	12	US-10-435-766-23	Sequence 23, Appl
44	86	3.6	850	10	US-09-738-626-5785	Sequence 5785, Ap
45	85.5	3.6	1185	10	US-09-895-913A-246	Sequence 246, App

ALIGNMENTS

RESULT 1

US-10-156-761-14253  
; Sequence 14253, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 14253  
; LENGTH: 454  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-14253

Query Match 32.6% Score 773.5; DB 15; Length 454;  
Best Local Similarity 38.7%; Pred. No. 7.8e-72;  
Matches 175; Conservative 81; Mismatches 183; Indels 13; Gaps 8;

Qy	6	DELNNLRNGKIDTVVLACVDMQRLMGKLTGRHFLGLDQKKISIST----
Db	11	BEELHVAAGGEIDTVLAFDPMQRLQGRKFAARFL---DEVLEHGTGCNYLLAVD 67
Qy	62	GIAGGGYEISVDYGSCHLCLADNLSHLLPWSGAVLAISNPHNFVTEPLFCSPRVI 121
Db	68	MNTVDGYDMSWRGDFAMHPDLSTLRQVWPNAGTALLIAD-LAWNDSPPVVAAPRQI 126

QY	122	LMQOIERLANLKLGLFASLEFNLFNFTYKASOKHWKNLKTAPHOMNITSASSGIE	181
Db	127	LRQLDRLAALGTAQVGTQTELEFVFKDTEYQAWDAGYKGLTPANQYNIDYSVLGTGRIE	186
QY	182	TFWBSVRNKLBEAGILMEATHPEFLPSQHELNFVPDPLTMADPHIITAKHGVEAEQSG	241
Db	187	PLURIRIRNEQAAGUTVDESAKGECPNGGQIEVFRYDEALVTCDOHAYKTKGAKEIAAQEG	246
QY	242	MYATFNAKLSSTALGNACHIIHMSLODAETEKAFYDQNDYEGMSTLARNMIAGLLKYVE	301
Db	247	VSTITFMAKYNERE-GNSCHIIHLSLADAD-GTNAMAGDGP-GMSDVMRHELAQLAALRD	303
QY	302	AYTFFASYINSYKRILOPIFTAPTCCWAIIDNRTSAFLCNKSKSEGINVELRICGADLNPY	361
Db	304	FSLLYAPNINSYKRFQPGSFAPTAWGYNDRNRTCALRV-CHGRSMRFENRLPGGVNPH	362
QY	362	LAFSAIIAAGISGIEKLELPPASGVYNDKELPEPPNSIQNATHLLKESKMLNKTFG	421
Db	363	LAVAGLVAAGLYGIEBHKLELPEACAGNAY-AAEYEHVPTTLREAAELWENSPTAKAAGD	421
QY	422	KUILHYVNAANVEINEFQSKQVTDWELNQGNR	453
Db	422	DVVAHYRNMARVELKAFDAVAATWELRRSFER	453

## RESULT 2

```

US-10-062-458-2
; Sequence 2, Application US/10062458
; Publication No. US20030003550A1
; GENERAL INFORMATION:
; APPLICANT: NAKAMURA, JUN
; APPLICANT: IZUI, HIROSHI
; APPLICANT: MORIGUCHI, KAYO
; APPLICANT: KAWASHIMA, HIROKI
; APPLICANT: NAKAMATSU, TSUYOSHI
; TITLE OF INVENTION: METHOD FOR PRODUCING L-GLUTAMINE BY FERMENTATION AND L-GLUTAMINE
; TITLE OF INVENTION: PRODUCING BACTERIUM
; FILE REFERENCE: 219181USO
; CURRENT APPLICATION NUMBER: US/10/062.458
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: JP 2001-28163
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: JP 2001-162806
; PRIOR FILING DATE: 2001-05-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Brevibacterium lactofermentum
; US-10-062-458-2

```

Query Match	20.0%	Score 475.5	DB 15	Length 446
Best Local Similarity	29.2%	Pred. No. 1.2e-40		
Matches 135	Conservative 83	Mismatches 174	Indels 71	Gaps 12

  

Qy	15	GKIDTVVLACVDMGRLMGKRLTGRHFLGLDOKKISITFYVYVTIEGIAGGVEISVD	74
Db	29	CHLKSVVVAELESAL-----EEGIGFD-----GSAIGYA-----RIEAD	66
Qy	75	TGYSDCHLCADLNSLHLPLWSEG-AVLAINSPHNFT---SEPLFCSPRVILMQOIERLA	130
Db	67	T-----IARDDPSTFQVLPLEAGISKLQARLFCVDVIMPDGQPSFSDPRQVLRQVQLAA	121
Qy	131	NLKLKGLFASELEFNLENFYKSASQXHKNLKT-----AQPHQMNNTSASSGIETEM	184
Db	122	DEGLTCMISPEIEFYLV-----QSLRTNGLPPVPTDNGGFDQATFNEANFR	169
Qy	185	RSVRNKLKEAGILMEATHPEFLPSQHELNFVPADPLTMADRHIIIAKHGVREMAEQSGVVA	244
Db	170	RNAMVALEEELIGIPVEFSGHHTAPQCOBIDLRHADLTMAADNIMTFYIMKVAADQGVGA	229

[illegible]

### RESULT 3

```

US-09-738-626-5942
; Sequence 5942, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKKO
; APPLICANT: SENO, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKAZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5942
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5942

```

	Query Match	19.8%; Score 470.5; DB 10; Length 446;
	Best Local Similarity	28.9%; Pred. No. 4.1e-40;
	Matches 134; Conservative	84; Mismatches 174; Indels 71; Gaps 12;
Qy	15 GKIDTVLVACVDMGRLMGKRLTGRHFLGLDQKKISISTFYAVYVTIEGIAGGYEISSVD	74
	: : : : : :	:
	: : : : : :	:
Db	29 GLHKSVVVPAELLESAL-----EEGIGFD-----GSAIEGYA---RISEAD	66
Qy	75 TGYSDCHLCADLNSHLHPWSEG-AVLAINPHNFVT---SEPILFCSPRVLMQIERLA	130
	: : : : : :	:
	: : : : : :	:
Db	67 T-----IARPDPTFOVLPLEAGISKLOARLFCDVTPMGPQSFSPPQVLRROVQLAA	121
Qy	131 NLKLKGIFASELGFNLFNETYKSASQKHWNKLT-----AQPHHWMNINSSASSGIETFM	184
	: : : : : :	:
	: : : : : :	:
Db	122 DEGLTCMSPIEIFYL-----QSLRTNGLPVPPTDNGGVFDQATFNAPNFR	169
Qy	185 RSVNKLKEEAGILMEATHPEPLPSOHELNVFPADPLTWADRHITAKHGVRMAEQSGNVA	244
	: : : : : :	:
	: : : : : :	:
Db	170 RNAMVALLEEIGIPVEFSHHETAQQCEIDURHADALTWAONIMTFRYMKVOVARDDQGGA	229
Qy	245 TFMAKLSSTALGNACTHMSLDQAETEKNAFYDONDEYGMSTLARNMIAGLLKLVPEATY	304
	: : : : : :	:
	: : : : : :	:
Db	230 SEMPKPPOEHAGSAMHTMSLFEGDT--NAFHDPDDSYMLSKTAQOFIAGTLIHHAPEFTA	287

Qy	305	FFASYINSYKRLQPIPTFAPTCKCWAIDNRTSAFLCN----	SKSGINVELRIGADLNPY	361
Db	288	VTNQWNSYKRIWYGNAPTAAATGVSNRSALVRVPTVRLNKESRRRVEVRLPDTACNPY	347	
Qy	362	LAFSAIIAAGISIGIEKLELPPPSAGV-----	YNOKELPEFFNSQNAHTLL	409
Db	348	LAFSVMLGAGLKGIEGVELDEPAEDDTISLNSFRERRAMGYND-----	LPSSLDQALRQW	402
Qy	410	KESKMLNKTTFGEKLILHYVNAANVEINFEFSQVTDWELNQGFN	452	
Db	403	EKSELVADILGHVFEFFRLKRWKREWRDYQOQITPEWLRNNLD	445	

RESULT 4  
US-09-815-242-13378  
; Sequence 13378, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes.

Query Match	18.2%;	Score 432;	DB 9;	Length 448;
Best Local Similarity	27.2%;	Pred. No. 4.4e-36;		
Matches	128;	Conservative 83;	Mismatches 22;	Indels 40; Gaps 10;
Qy	1	MTITYDELNNLRNGKIDTVVLACVDMOGRMGKRLTGTRHFLGDDO---	KKISISTFYVA	57
Db	1	MPITADIRREVKEKNTVFIEMSDILGTWKNVEIP-----	ATDEQLDKVLSKNKMPDG	55
Qy	58	VTIEGIAGGYEISSVDTGYSDCHLCADNLHLHPMSE--	GAVLAISNPNHFNVTSEPLF	115
Db	56	SSIEGF-----	VRINESDMYLPDLPTWTFVPPMGDENGSGVAGLICDVTYTEGEDPFA	106
Qy	116	CSRPVILMOQIERLANLKLKGL-FASELEFNLF--	NETYKSAQKHWNKLTQAHPHOM	172
Db	107	GDPRGNLKRALRHMEEVGKSFNLGPBPFFFLKLDNGDPTLVNDKKG-----	GYP	158
Qy	173	NISASSGIEFTMRSVRNKLEAGILMEATHPFLPQSHELNFVPADPLTMADRIIAKHG		232
Db	159	DLAPTDLADNTRRIIVNVLTKMGFEVSAHSEVAVGOHEIDFKYDEVLRACDKIOI	PKLV	218

	Qy	233	VREMAEQSMVATFMAKLSSTALGNACHIIHMSLODAETKNAFYDONDEYGM--STLARN	290
			: :   :         :	:
	Db	219	VKTIRKHGLYATFMARPKFGIAGSGMHCHNSLFDAE--GNNAFDPNDPKGQLSETAYH	277
			: :   :         :	:
	Qy	291	WIAGLLKYVEATVPFASYINSYKRIQPLTFAPTKCCWAIDNRSAFLCNHSKSGSINVTE	350
			: :   :   :   :         :	:
	Db	278	FLOGLIKHAYNYTAIMNPTVNSYKRLPVGEAPVYIAWAGRNRSLPVRPASRGMGTGLE	337
			: :   :   :   :         :	:
	Qy	351	LRIIGADNLPLYAPSATIAAGISGIEEKLELPSPASGNVY-----NDKELPEFPNSIQ	403
			: :   :   :   :         :	:
	Db	338	LRSVDPMANPYVAMAVLLEVLGYLGIENTKIAPAIEENIYYMTABERKEAGITDLPSTLH	397
			: :   :   :   :         :	:
	Qy	404	NATHLLKESKMNLNKFGEKLILHYVNVAANVEINFEFSKVQVTDWELNQGFNRY	454
			: :   :   :   :         :	:
	Db	398	NALKALTEDEVYKAALGDHIYTSFEAKRIEWSGVATFVSOWEIDNYLDLY	448
			: :   :   :   :         :	:

RESULT 5  
US-10-156-761-13481  
: Sequence 13481, Application US/10156761  
: Publication No. US20030119018A1  
: GENERAL INFORMATION:  
: APPLICANT: OMURA, SATOSHI  
: APPLICANT: IKEDA, HARUO  
: APPLICANT: ISHIKAWA, JUN  
: APPLICANT: HORIKAWA, HIROSHI  
: APPLICANT: SHIBA, TADAYOSHI  
: APPLICANT: SAKAKI, YOSHIYUKI  
: APPLICANT: HATTORI, MASAHIRA  
: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
: FILE REFERENCE: 249-262  
: CURRENT APPLICATION NUMBER: US/10/156,761  
: CURRENT FILING DATE: 2002-05-29  
: PRIOR APPLICATION NUMBER: JP 2001-204089  
: PRIOR FILING DATE: 2001-05-30  
: PRIOR APPLICATION NUMBER: JP 2001-272697  
: PRIOR FILING DATE: 2001-08-02  
: NUMBER OF SEQ ID NOS: 15109  
: SEQ ID NO 13481  
: LENGTH: 453  
: TYPE: PRT  
: ORGANISM: Streptomyces avermitilis  
US-10-156-761-13481

Query Match	18.0%; Score 427.5; DB 15; Length 453;
Best Local Similarity	30.0%; Pred. No. 1.3e-35;
Matches	127; Conservative 74; Mismatches 167; Indels 55; Gaps 12;
Qy	61 EGIAGGGEYESSVDYGY-SDCHLCADNLNSHLHLPWSEGA-----VLAINPHNFVTVSEP 113
Db	47 EGIGFGSAIEGFGARVYESDMIAKPPDSTQVLPFWRAEAFGTARMFCDILMP-----DGSP 101
Qy	114 LFCSPRVILMQOIERLANLKLGLFASELEFNLFNETYKSAQSGHKWNLKLTAAQ- - - - - 167
Db	103 SFADPRVYLKBAKAKTSDLGFTFYTHPEIEFFLLKQ-----KPLGSRRTPADNS 152
Qy	168 ---HHOMNISASGIETNRSVRNKLKEEAGILMEATHPEFLPSQHELNFVPADPLTWAD 224
Db	153 GYPDHTPQNV- ---GMD-FRRQAITMLSEMGISVEFSFHEGAPCQOQIDILRYADALSTAD 207
Qy	225 RHIIAKHGVRMAEQSGWATFWAKLSSATLGNACHTHMSLQDAETEKNAFYDQNDVEGVM 284
Db	208 NIMTFRLVMQVAALEQGVQATFMKPPSEYVPGSGMHTLSL- - - - -FEGDRNAPYESGSEYQL 265
Qy	285 STILARNWIAGLLKYVPEATYFFASINSYKRLQPLT-----FAPTKCCWAIDNRKTS 336
Db	266 SKVGRSFIAGLLKHAAEISAVTQWNSYKRIWGGSERTAGAGEAPSYICWGHNRSAL 325
Qy	337 FRL- ---CNSKSEGINVELRTGGADLANPYLAFSAIIAAGISGIEKLELPPPPASGNVY- - - 391
Db	326 VRVPMYKPGKTSARVEVRSLDGSAIPYLAAYALLAAGLKGISEGYLPLPGCADDVWALS 385



Qy 233 VREABOSGMVATFMAKLSSTALGNACHIHMSLODAETKNAFYDQNDQYGMSTLARNWI 292  
Db 220 VKTIARKHNLHATFMPKPLFGVNGSGMHFNVS--FKGENAFDPNTEMLGTETAYQFT 277  
Qy 293 AGLKYYPEATYFASYSYNSYKRLQPLTFAPTKCCWALDNRSTAFRLCNSKSEGINVELR 352  
Db 278 AGVLKARGFTAVCNPLVNSYKRLVPGYEAPCYIANGSKNSPLIRVPSRGLSTRLEVR 337  
Qy 353 IGGADLNPLAFSAIIAAGISGIEEKLPPPPASGNVY--NDKE-----LPEFPNSLQNA 405  
Db 338 SVDPAANPYMALAAILLEAGLDGINKLKVPPEVQNQIYENNREREAVGIQDLPLSTLYTA 397  
Qy 406 THLLKESKMLNKTGKELILHYNAANVEINEFSKQVTDWELNQGFRY 454  
Db 398 LKAMRENEVIKKGALNHIYNQFINSKSIWDYRTOVSEWERDQYMKQY 446

## RESULT 8

US-09-815-242-12775

; Sequence 12775, Application US/09815242  
; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR FILING DATE: 2000-03-21

; PRIOR FILING DATE: 2000-03-21

; PRIOR FILING DATE: 2000-05-23

; PRIOR FILING DATE: 2000-05-23

; PRIOR FILING DATE: 2000-05-23

; PRIOR FILING DATE: 2000-05-26

; PRIOR FILING DATE: 2000-05-26

; PRIOR FILING DATE: 2000-10-23

; PRIOR FILING DATE: 2000-10-23

; PRIOR FILING DATE: 2000-11-27

; PRIOR FILING DATE: 2000-11-27

; PRIOR FILING DATE: 2000-12-22

; PRIOR FILING DATE: 2000-12-22

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 12775

; LENGTH: 446

; TYPE: PRT

; ORGANISM: Staphylococcus aureus

US-09-815-242-12775

Query Match 17.9%; Score 424.5; DB 9; Length 446;  
Best Local Similarity 26.2%; Pred. No. 2.7e-35;  
Matches 123; Conservative 92; Mismatches 211; Indels 43; Gaps 9;

Qy 2 TITVDLNNLRNGKIDTVVACVDMQGRMLGKRLTGRHFLGLDQKKISISTFVYAVTIE 61  
Db 5 TTFKDDIRKFAEENVYRLRQFTDILGTIKNVEVPVSQL-----EKVLDMEMFDGSSIE 60  
Qy 62 GIAGGGVEISVDTGYSDCHLCADLNSLHLLPWS--EGAVLAINPNHNFVTSEPLFCSPR 119  
Db 61 GF-----VRIEESDMVLHPDLDTWIFPWTAGQGVKVARLICDVYKTDGTFEGDPR 111  
Qy 120 VILMQOIERLANLKLKGL-FASELEFNLFNETHYKASQKQKWNKLTAPHPHQ-----WM 172  
Db 112 ANLKRVLKEMEDLGFTDFNLGPEPEFFLKLDEK-----GEPTLELNDGGYF 159

Qy 173 NISASSGIETFMRSVRNKLBEAGILMEATHPEFLPSQHLELNFVADPLTWADRHIIAKHG 232  
Db 160 DLAPTDLGENCRDRIVLELSDMGFDIEASHHEVAPGQHEIDFKYADAVTACDNIQTOKLV 219  
Qy 233 VREABOSGMVATFMAKLSSTALGNACHIHMSLODAETKNAFYDQNDQYGMSTLARNWI 292  
Db 220 VKTIARKHNLHATFMPKPLFGVNGSGMHFNVS--FKGENAFDPNTEMLGTETAYQFT 277  
Qy 293 AGLKYYPEATYFASYSYNSYKRLQPLTFAPTKCCWALDNRSTAFRLCNSKSEGINVELR 352  
Db 278 AGVLKARGFTAVCNPLVNSYKRLVPGYEAPCYIANGSKNSPLIRVPSRGLSTRLEVR 337  
Qy 353 IGGADLNPLAFSAIIAAGISGIEEKLPPPPASGNVY--NDKE-----LPEFPNSLQNA 405  
Db 338 SVDPAANPYMALAAILLEAGLDGINKLKVPPEVQNQIYENNREREAVGIQDLPLSTLYTA 397  
Qy 406 THLLKESKMLNKTGKELILHYNAANVEINEFSKQVTDWELNQGFRY 454  
Db 398 LKAMRENEVIKKGALNHIYNQFINSKSIWDYRTOVSEWERDQYMKQY 446

## RESULT 9

US-09-815-242-10514

; Sequence 10514, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR FILING DATE: 2000-03-21

; PRIOR FILING DATE: 2000-03-21

; PRIOR FILING DATE: 2000-05-23

; PRIOR FILING DATE: 2000-05-23

; PRIOR FILING DATE: 2000-05-26

; PRIOR FILING DATE: 2000-05-26

; PRIOR FILING DATE: 2000-10-23

; PRIOR FILING DATE: 2000-10-23

; PRIOR FILING DATE: 2000-11-27

; PRIOR FILING DATE: 2000-11-27

; PRIOR FILING DATE: 2000-12-22

; PRIOR FILING DATE: 2000-12-22

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 10514

; LENGTH: 451

; TYPE: PRT

; ORGANISM: Enterococcus faecalis

US-09-815-242-10514

Query Match 16.9%; Score 402; DB 9; Length 451;  
Best Local Similarity 28.9%; Pred. No. 6.1e-13;  
Matches 114; Conservative 63; Mismatches 184; Indels 34; Gaps 7;

Qy 78 SDCHLCADLNSLHLLPWSG-----AVLAINPNHNFVTSEPLFCSPRVLMOQIERLAN 131  
Db 73 SDWLYLPDLSTWIFPWSHDGKVARLICDIYNP----DGTTPAGDPR-----GN 118  
Qy 132 LKUKGLFASLEFNLFN-----ETYSASQKQKWNKLTAPHPHQMNISSASSGIETPMRS 186  
Db 119 LKRALADMKELGFTSFNLGPEPEFFLKLDEGEITTDLNDKGGYFDFAPTDLGNCRRD 178

QY 187 VRNKEEAGILMEATHPELPSOHELNFVPADPLTMADRHIIAKHGVRMAEQSGMVATF 246  
 DB 179 IVLESLGFEVASHHAPGHEIDFKYADVVDACDNIQTFKLVKTIARKHGLHATF 238  
 QY 247 MAKLSSTALGNACHIIHMSLODAETKNAFYDQNDGYMSTLARNWITAGLLKXVPEATYFF 306  
 DB 239 MPKPLPGINGSGMHCNMSL--FNEEGNVFDSGEMGLSOTAYHFLGGLLKUARAYTAVC 296  
 QY 307 ASVINSYKLOPLTFAPTCKCWAIDNRTSAFLCNKSKSGINVELRIGGADLNPLAFSA 366  
 DB 297 NPTVNSYKRLVPGYEAPVYVAMSGNRSPILIRVPESRGLSTRLELRSDPSPANPLYTMVAV 356  
 QY 367 ITAAGISGIEEKLLEPPASGNVY--NDKE-----LPEPPNSIQNATHLLKESKMLNKTFF 419  
 DB 357 LIQAGLDGRNELTTPPADVRNIIYVNBEEQHAQIEDLPSTLHNAIKELRKQDWMIDAL 416  
 QY 420 GEKLILHYVNAANVEINEFSKVQTDWELNQGNRY 454  
 DB 417 GRHIFANFVEAKRMWAAFRQTVSEWEREQYLELY 451

## RESULT 10

US-09-738-626-6985

; Sequence 6985, Application US/09738626

; Publication No. US20020197605A1

; GENERAL INFORMATION:

; APPLICANT: NAKAGAWA, SATOSHI

; APPLICANT: MIZOGUCHI, HIROSHI

; APPLICANT: ANDO, SEIKO

; APPLICANT: HAYASHI, MIKIRO

; APPLICANT: OCHIAI, KEIKO

; APPLICANT: YOKOI, HARUHIKO

; APPLICANT: TATEISHI, NAKO

; APPLICANT: SENOH, AKIHIRO

; APPLICANT: IKEDA, MASATO

; APPLICANT: OZAKI, AKIO

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-125

; CURRENT APPLICATION NUMBER: US/09/738,626

; CURRENT FILING DATE: 2000-12-18

; PRIOR APPLICATION NUMBER: JP 99/377484

; PRIOR FILING DATE: 1999-12-16

; PRIOR APPLICATION NUMBER: JP 00/159162

; PRIOR FILING DATE: 2000-04-07

; PRIOR APPLICATION NUMBER: JP 00/280988

; PRIOR FILING DATE: 2000-08-03

; NUMBER OF SEQ ID NOS: 7059

; SOFTWARE: PatentIn ver. 3.0

; SEQ ID NO 6985

; LENGTH: 477

; TYPE: PRT

; ORGANISM: Corynebacterium glutamicum

US-09-738-626-6985

Query Match

Best Local Similarity 13.6%; Score 322; DB 10; Length 477;

Matches 124; Conservative 74; Mismatches 207; Indels 62; Gaps 14;

QY 4 TYDELNNLRNGKIDTVVLACVDMQGRMLGKRLTGRHFLGLDQKKISISTFVYAVTIEGI 63

DB 5 TPEIVKFIKDNVEFVDRFTDLPG-----TEQHF-----SIPAAAFDADTIEGL 51

QY 64 AGGGYISSVDTG--GYSDCHLCADLNSHLPLWSEGAVLAIS--NPHNFVTSPLFCSPRVI 121

DB 52 AFDGSSIRGFTTIDSDMNLPLDGTATLDPFRKAKTLNVKFFVHDPPFTREAFSRDPRNV 111

QY 122 LMQOIERLANLKUGL--PASELEFNLNET-----YKSASQKHWNKLTQAQPH 169

DB 112 ARKAEQYLASTGADTCNFGAEAEFYLFDSVRYSTEMNSGFYVDTEEGWNRGKTNLD 171

QY 170 QMWNISASSGIE-----TFMRS--VRNKEEAGILMEATHPELPSQHELNFV 215

DB 172 GTNLAGKRNKVGKGYFFVAPYDQTVDRDDMVNRN--LAASGAFALRPFHFHVGQGCQINVR 230

QY 216 PADPLTMADRHIIAKHGVRMAEQSGMVATFMAKLSSTALGNACHIIHMSLODAETKNAF 275  
 DB 231 FNTMLHAADDQTQFKIINKVNTARLHGKAATFMPKPLAGDNGSGMHAQSLW--KDGKPLP 288  
 QY 276 YDQNDGYMSTLARNWITAGLLKXVPEATYFFASYINSYKRLQLOPLTFAPTCKCWAIDNRTS 335  
 DB 289 HDESGVAGLSDIARYIIGILHHAGAVLAFTNATLNSYHRLVPGPEAPNLVYSQRNSA 348  
 QY 336 APR--LCNKSSEGINVELRIGGADLNPLAFSAIIAAGISGIEEKLLEPPASGNVY--- 390  
 DB 349 AVRIPITGSPKAKRIEFPADPSGNPYLGFPAAMMAGLDGKRIEHPHAPVDKOLYELP 408  
 QY 391 --NDKLEPEFNS-----LQNATHLLKESKMLNKTFFGEKLILHYV 428  
 DB 409 PEEAASIFQAPTSLASLKAQEDTDFLTESDV-----FTEDLIEAYI 451

## RESULT 11

US-10-156-761-13532

; Sequence 13532, Application US/10156761

; Publication No. US20030119018A1

; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI

; APPLICANT: IKEDA, HARUO

; APPLICANT: ISHIKAWA, JUN

; APPLICANT: HORIKAWA, HIROSHI

; APPLICANT: SHIBA, TADAYOSHI

; APPLICANT: SAKAKI, YOSHIYUKI

; APPLICANT: HATTORI, NASHIRA

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-262

; CURRENT APPLICATION NUMBER: US/10/156,761

; CURRENT FILING DATE: 2002-05-29

; PRIOR APPLICATION NUMBER: JP 2001-204089

; PRIOR FILING DATE: 2001-05-30

; PRIOR APPLICATION NUMBER: JP 2001-272697

; PRIOR FILING DATE: 2001-08-02

; NUMBER OF SEQ ID NOS: 15109

; SEQ ID NO 13532

; TYPE: PRT

; ORGANISM: Streptomyces avermitilis

US-10-156-761-13532

Query Match

Best Local Similarity 12.2%; Score 291; DB 15; Length 469;

Matches 113; Conservative 68; Mismatches 189; Indels 66; Gaps 13;

QY 6 DELNNLRNGKIDTVVLACVDMQGRMLGKRLTGRHFLGLDQKKISISTFVYAVTIEGIAG 65

DB 6 DEAKYIADDEVKFDVVRFCDDLPQVY-----QHF-----TIPAEAFDPA---EELAF 49

QY 66 GGYEISSVDTGY--SDCHLCADLNSHLPLWSEGAVLAIS--NPHNFVTSPLFCSPRVI 123

DB 50 DGGSSIRGQAIHESDMALRADUSTARVDFRDKTVNINFFIHDPTIGQYSRDPENVAK 109

QY 124 QOIERLANKUL--KGLFASELEFNLNET--YKSASQKHWNKLT----- 164

DB 110 KAEAYLASTGIADTAYFGPEAEFYVDSVRFKTAENESFYHIDSEAGAWNTGAIEDNRY 169

QY 165 -----AOPHHQWNISASSGIEITFMRSVRNKEEAGILMEATHPEF--LPSQHELN 213

DB 170 KVRVKGGYFPTPPVDFHFDLRAEISLE-----LANSGLQVERQHHEVGTAGAEIN 220

QY 214 FVPADPLTMADRHIIAKHGVRMAEQSGMVATFMAKLSSTALGNACHIIHMSLODAETEN 273

DB 221 YKFNTLAAADDQLFKYIVKVNVRNGKTAFTMPKPIFGDNGSGMHHQSLW--ANGDP 278

QY 274 AYDQNDGYMSTLARNWITAGLLKXVPEATYFFASYINSYKRLQLOPLTFAPTCKCWAIDNR 333

DB 279 LFYDEAGYAGLSDTARYIYGILKHAQSLAFTNPTVNSYHRLVPGFEAPVNMVYSQRNR 338



```
QY 449 QGFNRY 454
Db 495 EYFEAF 500

RESULT 14
US-09-815-242-12088
; Sequence 12088, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12088
; LENGTH: 469
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-12088

Query Match 11.2%; Score 265.5; DB 9; Length 469;
Best Local Similarity 24.8%; Pred. No. 1.3e-18;
Matches 119; Conservative 79; Mismatches 194; Indels 87; Gaps 20;

QY 9 NNLRNGKIDTVVVLACVDMQGRMLGKRLTGRHFLGLDQKKISITFTFYAVTIEGIAGGY 68
Db 6 HQLIKDHVDKVVLDLRTDTYKGQOQVTPARD--ALDDREFF-----AGKWF 50

QY 69 EISSVDGTGSDCHLCADLNSLHLLPWSGAVLAISNPHNFVTSEPLFCSPRVIL----- 122
Db 51 DGSSI-AGWKG-----IEASDMILMPDDSTAVL-----DPTSEPTLLVCDIIE 94

QY 123 ---MQQIER---LAN-----LKLGL-----FASLEFNLNFNET-YKS-----A 154
Db 95 PSTMQGYERDPRNIAKRAEYLKSTGIDTVFGVPEPEFFIDFVFKSDISGSMFKIFS 154

QY 155 SQKHVK--NLKTAQPHQ-----WNLSASSGIETFMRSVKNKLEEAAGILMEATHPEF 205
Db 155 EQASWNTDIESGNKGRPGVKGYPFPVVDHDEHIRTAMCNALKEEGLVVEVHHVEV 214

QY 206 LPS-QHELVFPADPLTMADRHIIAKHGVREMAEQSGWATFMAKLSSTALGNACHIMS 264
Db 215 ATAQONEIGVKFTNLVAKADEVQTLKCVHNVDAYGKTVTFPKPLYGDNDSGMVHMS 274

QY 265 LQDAETEKAFYQNDYGNMSTLARNWIAGLLKYVPEATYFFASYINSYKRLQPLTFAPT 324
Db 275 I--SKDGKNTFAGEG-YAGLSETALYFIGIILKHGKALNGFTNPSTNSYKRLVPGPEAPV 331

QY 325 KCCWAIDNRTSAFRL-CNSKSEGINVELRIGGADLNPYLAFSAIIAAGISGIEEKLPP 383
Db 332 MLAYSARNRSASIRIPYVUSPKARRIEARFPDPAANPYLAFAPALMAGLDGIONKIHPGD 391

QY 384 PASGNVYN-----DKELPEFPNSLQATHLLKESK-MLNK--TFGEKLIHLHYVNAANVE 434
Db 392 AADKNLYDLPPBEAKEIPQVCGSLKEALELOKGRAFLTKGGVFTDEFIDAYIELKSEE 450

RESULT 15
US-09-815-242-11507
; Sequence 11507, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11507
; LENGTH: 481
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-815-242-11507

Query Match 11.2%; Score 265; DB 9; Length 481;
Best Local Similarity 24.7%; Pred. No. 1.5e-18;
Matches 100; Conservative 80; Mismatches 181; Indels 44; Gaps 13;

QY 77 YSDCHLCADLNSLHLLPWS-EGAVLAISNPHNFVTSEPLFCSPRVILMQOIRLANLKLK 135
Db 72 HSDMLTDPDLVRYFIDPFSDVSAVVCDVYDVYKQPYEKCPCRSIAKALKALHRLDLGLG 131

QY 136 GL--FASLEFNLNFNE-TYKSAS-----QKHWNKLTQA-----PHHQ--WM 172
Db 132 DVAYFCAENEFIFDPSIKIKDASNSQYVEVDSEGEWNDRKSFNGVNFHPRGQGGYM 191

QY 173 NTSASSGIEFTMRSVRNKLKEEAAGILMEATHPEFLPSQHELVFPADPLTMADRHIIAKHG 232
Db 192 PVPPTDPTMDIDITEIVKVLNVGLFTFVVHVEVAQAQGEVGVKFGDLVEAADNVOKLKIV 251

QY 233 VREMAEQSGWATFMAKLSSTALGNACHIMSLQDAETEKNAFYQNDYGNMSTLARNWI 292
Db 252 VQWVAHLNGKTAFTMPKPLYGDNDSGMVHVSIGW--KNNENLFGGETYK-GLSEFALYEL 308

QY 293 AGLLYKVPATYFFASYINSYKRLQPLTFAPTKCCWAIDNRTSAFRLCNKSE-GINVEL 351
Db 309 GGVLRHARGLAFTNASTNSYKRLIPGYEAPSIILTYASNSRNSASVRIPYGISKNARFEF 368
```





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 22:43:05 ; Search time 19 Seconds  
(without alignments)  
1011.007 Million cell updates/sec

Title: US-10-098-602A-2

Perfect score: 2376

Sequence: 1 MTITYDELNLIRNGKIDTV.....INEFSKQVTDWELNGFNRY 454

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCFUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	481	20.2	449	4	US-09-252-991A-17901
2	459	19.3	464	4	US-09-252-991A-27559
3	455	19.1	464	4	US-09-252-991A-28048
4	450	18.9	489	4	US-09-252-991A-27833
5	409	17.2	452	4	US-09-134-001C-4173
6	389	16.4	531	4	US-09-252-991A-26690
7	344.5	14.5	491	4	US-09-252-991A-19014
8	331.5	14.0	480	4	US-09-328-352-6949
9	317	13.3	1037	4	US-09-252-991A-16796
10	280.5	11.8	661	4	US-09-252-991A-18960
11	265.5	11.2	481	4	US-09-252-991A-32811
12	255.5	10.8	489	4	US-09-328-352-4363
13	241.5	10.2	190	4	US-09-107-532A-4613
14	134	5.6	275	4	US-09-107-532A-6200
15	95	4.0	481	4	US-09-328-352-4580
16	95	4.0	1338	1	US-08-471-033-50
17	95	4.0	1338	2	US-08-471-044-50
18	95	4.0	1338	2	US-08-463-483A-50
19	95	4.0	1338	2	US-08-471-046A-50
20	95	4.0	1338	2	US-08-470-566B-50
21	95	4.0	1338	2	US-08-469-334-50
22	95	4.0	1338	3	US-09-300-529-50
23	94	4.0	291	2	US-08-701-191A-24
24	94	4.0	606	3	US-08-335-865J-20
25	92	3.9	527	3	US-09-028-366-2
26	92	3.9	527	3	US-09-028-366-3
27	91.5	3.9	822	4	US-09-252-991A-29534

28	91	3.8	461	4	US-09-134-001C-5311	Sequence 5311, Ap
29	91	3.8	593	3	US-08-335-865J-21	Sequence 21, Appl
30	91	3.8	1013	3	US-09-415-522-8	Sequence 8, Appli
31	90	3.8	566	3	US-08-335-865J-9	Sequence 9, Appli
32	90	3.8	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
33	89.5	3.8	506	4	US-09-134-001C-4383	Sequence 4383, Ap
34	89.5	3.8	952	4	US-09-328-352-4551	Sequence 4551, Ap
35	88.5	3.7	731	4	US-09-252-991A-18769	Sequence 18769, A
36	88	3.7	2152	3	US-09-036-987A-3	Sequence 3, Appli
37	88	3.7	2152	3	US-09-370-700-3	Sequence 3, Appli
38	88	3.7	2152	4	US-09-603-207-3	Sequence 3, Appli
39	87.5	3.7	441	4	US-09-354-776B-79	Sequence 79, Appl
40	87	3.7	299	4	US-09-328-352-7993	Sequence 7993, Ap
41	87	3.7	733	4	US-09-328-352-5599	Sequence 5599, Ap
42	87	3.7	1074	1	US-07-912-015-4	Sequence 4, Appli
43	86.5	3.6	605	3	US-09-063-950-5	Sequence 5, Appli
44	86.5	3.6	3290	4	US-09-328-352-5486	Sequence 5486, Ap
45	86	3.6	973	4	US-09-107-532A-4810	Sequence 4810, Ap

ALIGNMENTS

RESULT 1  
US-09-252-991A-17901  
; Sequence 17901, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 17901  
; LENGTH: 449  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-17901

Query Match	20.2%	Score 481	DB 4	Length 449
Best Local Similarity	28.9%	Pred. No. 2.2e-44		
Matches	128	Conservative 94	Mismatches 181	Indels 40
Gaps	11			
QY	22	LACV--DMOGRMLGKRLTGRHFLGLDQKISITGFVAVVTIEGIAGGYEISSVDTGYSD	79	
Db	21	IECVTPDLNGVPRGKWTGEGF--LSGRRLQLAR---GVLLQCIIMGYPFARFVGSDDG	75	
QY	80	CHLCADLNSLHLPSWSEGAVALSNPHNFVTSE-----PLFCSPRVILMQOIERLANL	132	
Db	76	LALVAEPTQVHRLPWS-----NTPRAPAIQAQELDGTGSLSTRGLLKQVVARAAH	128	
QY	133	KLKGLFASELEFNLFF-----NETYKSASQKHWNKLTQAPHHQWNISASSGIETPMRS	186	
Db	129	GWQFVVAATELEFFVFPAPNTDPNEFPQAPLPGDGR---ELGYSAFSVSSNGLRPPFD	184	
QY	187	VRNKLBEAGFLMEATPEPLPSQHELNFFVADPLTWMADRHIIAKHGVRMAEQSGMVAIF	246	
Db	185	VYRCMDALGLVTRTFHEMGTSOFENFLHGDVPVLADQTFPLKLLKEVALKHGLIIVC	244	
QY	247	MAKLSSTALGNACHIIHMSLQDAETEKNAFYDQNDGYGMSTLARNWLAGLLKYVEATYFF	306	
Db	245	MAKPLAKTPGSSWHIHSQIVELDGGRNIFSEADGE--PSAAFRHFICGQACIADFTPF	302	
QY	307	ASTVNSYKRLQPLTFAPTKCCWADNRTSAFRCLNKSSEGINVELRTGGADLNPYLAFA	366	
Db	303	APHVNSFORLCHPYASPNNACWNSHDNRAAGLRIPASCASARRVENRUPGADANPYLAIA	362	





Db 412 I---PQGAARVECKAADICGPNYPYLAAGLEGIRKLDQPGQPHREMYHYSREVA 468  
Qy 395 ---LPEFPNSLQWATHLLKESKMLNKTGKELILHYVNAANVEINEFSSKQVTDWELNQ 449  
Db 469 QMGIELTPRLTSEDAIDAFADPLSRQVFGDAMYQAFVFKRDEWNAHYHTHVSDEWIEQR 526

## RESULT 7

US-09-252-991A-19014  
; Sequence 19014, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 19014  
; LENGTH: 491  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-19014

Query Match 14.5%; Score 344.5; DB 4; Length 491;  
Best Local Similarity 26.8%; Pred. No. 4e-29;  
Matches 122; Conservative 73; Mismatches 231; Indels 29; Gaps 11;  
Qy 11 LIRNGKIDTVVACVDMQGMKRLTGRHFLGLDQKKISISTFVAVTIEGAGGYEI 70  
Db 49 LAAPHEVRSITELFLIDANGVPRGKLLHRLDELLAVESGRPLPSTILGLTIN---GDDVEE 105  
Qy 71 SSV--DTGVSDCHLCADLNSLHLLPWSE---GAVLAISNPHNFTVSEPLFCSPRVILMQ 125  
Db 106 TGLVWDVGADCTWTFPLPGSLTLQPRQSTGQVQVSMHPELGLPAAA--GDRHVLQRV 163  
Qy 126 IERLANLKLGLFASLEFNLNFTYKS-----ASQKHWNKLTQAOPHHQWNNISASSG 179  
Db 164 IHSLOAEGHPVNAVELEFLLDRDSDGRPLPALQMGQPRATQVY---GVYLEQ 219  
Qy 180 IETFMRSVRNKLBEAGILMEATHPEPLPSQHELNFPVA-DPLTMADRHIIAKGVREMAE 238  
Db 220 LQPFLLDLVAAACEAQLPARTAISVAPGQVEITLQHRFDTLQAIDEGVRYKRLVKGVAN 279  
Qy 239 QSGWATFWAKLSSTALGNACHTHMSLQDAETEKNAFYQNDYGMSTLARNWIAGLLKY 298  
Db 280 RHGLQACFWAKPPADLSGGLHUVSLADA--AGNNLFASEDPAQ-TPLLRQAIGGMKAC 336  
Qy 299 VPEATYFFASYINRYKRLQPLTFAPTKCCWAIDNRTSAFRLCNSKSEGINVELRIGCADL 358  
Db 337 LLESALFCPNANSPRFOANSVAPLPTWGINNRTVSLRVPGPASSRHIEHICGADA 396  
Qy 359 NPYLAFSAIAAGISGIEEKLPPPPASGNVYNDKELPEFPNSLQWATHLLKESKMLNKT 418  
Db 397 NPYLAALAAALVRLGIRERLDPGAITNGY-AQATOALPSDWLTALRALEGSAAWAREA 455  
Qy 419 FGEKLILHYVNAANVEINEFSSKQVTD-----WELNQ 449  
Db 456 LGEDFLKILYLAIKQAEYRAPMGVEGQDWRWYLNQ 490

## RESULT 8

US-09-328-352-6949  
; Sequence 6949, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 6949  
; LENGTH: 480  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-6949

Query Match 14.0%; Score 331.5; DB 4; Length 480;  
Best Local Similarity 25.7%; Pred. No. 1.1e-27;  
Matches 117; Conservative 76; Mismatches 187; Indels 75; Gaps 14;  
Qy 26 DMQRLMGKRLTGRHFLGLDQKKISISTF---VTAVTIEG-----IAGGYEISSVDT-- 75  
Db 68 DLNGHIRGKRIDVKSL-----KNLSNGCYFPLSVYVAMSLDGKVIETGLGKYIGEPDRLC 122  
Qy 76 -----GYSDCHLCADLNSLHLLPWSEGAVALAISNPHNFTVSEPLFC--SPRVILMQQIER 128  
Db 123 LPILGSLQPSALSPELNAQLYLSMQE-----EDSGCRYEPRNLIKLLNQ 168  
Qy 129 LANLKLKGLFASLEFNLNFTYKSASQKHWNKLTQAOPHHQWNNISASSGIETFMRSVR 188  
Db 169 LHANNYFPVMAAELEFYLFSPOHQS-----EISENOCFDIDAPNNYQ----QVL 214  
Qy 189 NKLEBAGIL---MEATHPEPLPSQHELNFPV-PADPLTMADRHIIAKGVREMAEBSGMV 243  
Db 215 DEVERAALLQSTIEITAIVAESSPGQVELNLQNHDLILKCDQINALKRIVKQVARKHDLT 274  
Qy 244 ATFMKLSSTALGNACHTHMSLQDAETEKNAFYQNDYGMSTLARNWIAGLLKYVPEAT 303  
Db 275 ACFMAKPNLAKASGGMHFMSLN-QYHONIFSSBEKDEKALLKLSAISGLIELMPASM 333  
Qy 304 YPFASYINRYKRLQPLTFAPTKCCWAIDNRTSAFRLCNSKSEGINVELRIGGADLNPYLA 363  
Db 334 ALAPNINSYRRFKIGHVPLEANWDTNRRNVAIRIPCSDVQNRLEYRVAGADCNPLYLV 393  
Qy 364 PSATIAAGISGIEEKLPPPPASGNVYNDKELPEFP-----NSLQWATHLLKESKMLN 416  
Db 394 TATILAGASVGLSHKLLPLPKA-----HLLKFPDEHILLANNQPEALKIFKGSLLIK 445  
Qy 417 KTFGEKLILHYVNAANVEINEFSSKQVTD-----WEL 447  
Db 446 GYLGAADFVEHWTYVQAEYQNIYSQMTAEABQHWDI 480

## RESULT 9

US-09-252-991A-16796  
; Sequence 16796, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 16796  
; LENGTH: 1037  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (769)  
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.



```
Db 398 AADKNLYLPPPEAKBIQVCGSLKEALBELDKGRAFTKGGVFTDEFIDAYIELKSEE 456

RESULT 12
US-09-328-352-4363
; Sequence 4363, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4363
; LENGTH: 489
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4363

Query Match 10.8%; Score 255.5; DB 4; Length 489;
Best Local Similarity 24.7%; Pred. No. 3.1e-19;
Matches 98; Conservative 57; Mismatches 156; Indels 85; Gaps 12;

QY 112 EPLFCSPRVLIM-----QQIERLANLKLK-----GLFASLEFNLFN 148
Db 97 DPFPAEPTVVVTCVIEPSTGCGYERDRPSIARAEEYLKSTGICGTATFGPEPEFFVD 156
QY 149 ET-----YKSASQKHWKNLK-----TAQPHHQMWN 173
Db 157 EVKWDIDMSGARHTLIAEBAWSTGKDYESGNSGHRPVKGGYFPVPVDSAQD----MR 212
QY 174 ISASSGIEFMRSVRKNLEAGILMEATHEPFLPSQHELNFPVADPLTWADRHIIAKGV 233
Db 213 AEMCAKIEDIMGPR-----VEVHHHEVASQLEIGVSFNTLVKRADEVQQPKYAV 263
QY 234 REMAEGSGVATPMAXLSTALGNACHIHMSLQDAETEKNAFYDQNDDEV-GMSTLARNMI 292
Db 264 WNVHQQYATATMPKPMQVNDGNGMHVHMSI---SKDGKLP--AGDEVAGISEMALYFI 319
QY 293 AGLLKYVPEATYFFASYINSYKRLQPLTFAPTCKCWAIDNRITSAPRL-CNSKSEGINVEL 351
Db 320 GGIHKHARALNATNPSTNSYKELVPHFEAPIMLAYSAARNRSASIRIPVSNPKGRKIEA 379
QY 352 RIGGADLNPLAFSAIIAGISGIEKLELPPPPASGNVY-----DKELPEFPNSLQNAI 406
Db 380 RPDPMNPNYGLPAALLMAGIDGIONKIHFGAADKNLYDLPPPEEAKIPTVAHSLDMAL 439
QY 407 HLK---ESKMLNKTGKILHYVNAANVEINEFS 439
Db 440 EALQADHFLKGGVTKEMLDAYIELKTEDVRRLN 475

RESULT 13
US-09-107-532A-4613
; Sequence 4613, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC

Db 398 AADKNLYLPPPEAKBIQVCGSLKEALBELDKGRAFTKGGVFTDEFIDAYIELKSEE 456

OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 4613:
SEQUENCE CHARACTERISTICS:
LENGTH: 190 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc_feature
LOCATION: (B) LOCATION 1...190
SEQUENCE DESCRIPTION: SEQ ID NO: 4613:
US-09-107-532A-4613

Query Match 10.2%; Score 241.5; DB 4; Length 190;
Best Local Similarity 29.5%; Pred. No. 2.1e-18;
Matches 56; Conservative 38; Mismatches 89; Indels 7; Gaps 2;

QY 272 KNAFYDQNDGYGMSLTARNIAGLLKYVPEATYFFASYINSYKRLQPLTFAPTCKCWAID 331
Db 1 KKCLYDKGGELESETAYHFLGGLLKARAYTAVCNPTVNSYKRLVPGVEAPVYVWMSGR 60
QY 332 NRTSAFRLCNSKSEGINVELRIGGADLNPLAFSAIIAGISGIEKLELPPPPASGNVY- 390
Db 61 NRSPLVRVPESRGLSTRLESLRSDPSANPYLAMAVLQAGLDGIRNQITPPPAVDNRNIYV 120
QY 391 NDKE-----LPEFPNSLQNAIHLKESKMLNKTGKILHYVNAANVEINEFSKQVTD 444
Db 121 MDEEREAHIOQLPSTIHNAIKELRKONVMIDALGQHIFSVEAKLEWAAFRQTVSE 180
QY 445 MELNQGFNRY 454
Db 181 WEREQYLELY 190

RESULT 14
US-09-107-532A-6200
; Sequence 6200, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
```





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 22:37:15 ; Search time 39 Seconds  
(without alignments)  
1847.740 Million cell updates/sec

Title: US-10-098-602A-2

Perfect score: 2376

Sequence: 1 MTITYDLNLLIRNGKIDTV.....INBFKQVTDWELNQGFNRY 454

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- A Geneseq\_19Jun03.\*
- 1: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1980.DAT.\*
  - 2: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1981.DAT.\*
  - 3: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1982.DAT.\*
  - 4: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1983.DAT.\*
  - 5: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1984.DAT.\*
  - 6: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1985.DAT.\*
  - 7: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1986.DAT.\*
  - 8: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1987.DAT.\*
  - 9: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1988.DAT.\*
  - 10: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1989.DAT.\*
  - 11: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1990.DAT.\*
  - 12: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1991.DAT.\*
  - 13: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1992.DAT.\*
  - 14: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1993.DAT.\*
  - 15: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1994.DAT.\*
  - 16: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1995.DAT.\*
  - 17: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1996.DAT.\*
  - 18: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1997.DAT.\*
  - 19: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1998.DAT.\*
  - 20: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1999.DAT.\*
  - 21: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2000.DAT.\*
  - 22: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2001.DAT.\*
  - 23: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2002.DAT.\*
  - 24: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	614.5	25.9	459	22 AAG80030	Pseudomonas sp KIE
2	614.5	25.9	459	22 AAG80038	Pseudomonas putida
3	502.5	21.1	472	22 AAU54831	Propionibacterium
4	475.5	20.0	446	22 ABP3500	Brevibacterium lac
5	470.5	19.8	446	22 AAG92188	C glutamicum prote
6	470.5	19.8	446	22 ABP79681	Corynebacterium gl
7	464	19.5	446	23 ABBS5555	Lactococcus lactis
8	453	19.1	445	23 ABP66039	Bifidobacterium lo
9	445	18.7	448	23 ABP27862	Streptococcus poly

10	441	18.6	448	23 ABP27863	Streptococcus poly
11	432	18.2	448	22 AAU37785	Streptococcus pneu
12	432	18.2	448	24 AAE33506	Streptococcus pneu
13	432	18.2	448	24 ABU00865	S. pneumoniae type
14	424.5	17.9	446	22 AAU34237	Staphylococcus aur
15	424.5	17.9	446	22 AAU36745	Staphylococcus aur
16	424.5	17.9	446	22 AAU37182	Staphylococcus aur
17	410	17.3	444	23 ABB48084	Listeria monocytog
18	409	17.2	446	22 AAG82578	S. epidermidis ope
19	409	17.2	446	22 ABP39328	S. epidermidis ope
20	402	16.9	451	22 AAU34921	Enterococcus faeca
21	400.5	16.9	443	10 AAF93191	Peptide with glut
22	393.5	16.6	446	19 AAW37905	Mycobacterium tube
23	389.5	16.4	483	22 AAB96475	Putative P. abyss
24	333	14.0	845	23 ABB92646	Herbicidally activ
25	322	13.6	477	22 AAG93231	C glutamicum prote
26	322	13.6	477	22 AAB79882	Corynebacterium gl
27	306	12.9	478	18 AAW18160	Mycobacterium tube
28	305.5	12.9	478	23 ABP65893	Bifidobacterium lo
29	303	12.8	478	19 AAW63035	Mycobacterium tube
30	303	12.8	478	19 AAW37901	Mycobacterium tube
31	294.5	12.4	533	22 AAU54574	Propionibacterium
32	287	12.1	472	22 AAU35544	Haemophilus influe
33	287	12.1	1038	22 ABG24868	Novel human diagno
34	281.5	11.8	500	24 ABP57636	S. murayamaensis A
35	265.5	11.2	469	22 AAU36495	Pseudomonas aerugi
36	265	11.2	481	22 AAU35914	Helicobacter pylor
37	255.5	10.8	524	24 ABP80864	N. gonorrhoeae ami
38	247	10.4	469	22 AAU38196	Salmonella typhi c
39	241	10.1	469	22 AAU34825	E. coli cellular p
40	216.5	9.1	816	22 ABG13707	Novel human diagno
41	201.5	8.5	476	22 ABG25229	Novel human diagno
42	180.5	7.6	185	19 AAY85879	S. pneumoniae deri
43	169.5	7.1	334	22 AAU23182	Novel human enzyme
44	160.5	6.8	118	23 ABP32638	Human ORF1611 prot
45	147.5	6.2	159	21 AAB41477	Human ORFX ORF1241

ALIGNMENTS

RESULT 1

AAAG80030  
ID AAAG80030 standard; Protein; 459 AA.  
XX  
AC AAG80030;  
XX  
DT 17-JAN-2002 (first entry)  
XX  
DB Pseudomonas sp KIE171-BII ipuC protein.  
XX  
KW L-alaninol; isopropylamine; iup1; ipuH; ofloxacin; ipuC; glutamylamide;  
KW gamma-glutamylamide synthase; theanine.  
XX  
OS Pseudomonas sp.  
XX  
PN WO200173038-A2.  
XX  
PD 04-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-EF03651.  
XX  
PR 31-MAR-2000; 2000EP-0106888.  
XX  
FA (LONZ ) LONZA AG.  
XX  
PI Leisinger T, Van Der Ploeg J, Kiener AM, De Azevedo Waesch SI;  
PI Maire T;  
XX  
DR WPI; 2001-626266/72.  
DR N-PSDE; AAI68686.  
XX  
PT New microorganism for converting isopropylamine to L-alaninol, useful



QY 52 STFVAVTIEG--IAGGGEISSVDTGY-----SDCHLCADLNSLHLLPWSEGAVLA--- 101



AC AAB79681;  
 XX 30-APR-2001 (first entry)  
 DT  
 XX  
 DE Corynebacterium glutamicum MP protein sequence SEQ ID NO:96.  
 DE  
 XX  
 KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;  
 KW fine chemical production; microorganism; organic acid; nucleoside;  
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;  
 KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;  
 KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN WO200100843-A2.  
 XX  
 XX 04-JAN-2001.  
 PD  
 XX  
 PF 23-JUN-2000; 2000WO-IB00923.  
 XX  
 XX 25-JUN-1999; 99US-0141031.  
 PR 01-JUL-1999; 99DE-1030476.  
 PR 02-JUL-1999; 99US-0142101.  
 PR 08-JUL-1999; 99DE-1031415.  
 PR 08-JUL-1999; 99DE-1031418.  
 PR 08-JUL-1999; 99DE-1031419.  
 PR 08-JUL-1999; 99DE-1031420.  
 PR 08-JUL-1999; 99DE-1031424.  
 PR 08-JUL-1999; 99DE-1031428.  
 PR 08-JUL-1999; 99DE-1031434.  
 PR 08-JUL-1999; 99DE-1031435.  
 PR 08-JUL-1999; 99DE-1031443.  
 PR 08-JUL-1999; 99DE-1031453.  
 PR 08-JUL-1999; 99DE-1031457.  
 PR 08-JUL-1999; 99DE-1031465.  
 PR 08-JUL-1999; 99DE-1031478.  
 PR 08-JUL-1999; 99DE-1031510.  
 PR 08-JUL-1999; 99DE-1031541.  
 PR 08-JUL-1999; 99DE-1031573.  
 PR 08-JUL-1999; 99DE-1031592.  
 PR 08-JUL-1999; 99DE-1031632.  
 PR 08-JUL-1999; 99DE-1031634.  
 PR 08-JUL-1999; 99DE-1031636.  
 PR 09-JUL-1999; 99DE-1032125.  
 PR 09-JUL-1999; 99DE-1032126.  
 PR 09-JUL-1999; 99DE-1032130.  
 PR 09-JUL-1999; 99DE-1032186.  
 PR 09-JUL-1999; 99DE-1032206.  
 PR 09-JUL-1999; 99DE-1032227.  
 PR 09-JUL-1999; 99DE-1032228.  
 PR 09-JUL-1999; 99DE-1032230.  
 PR 14-JUL-1999; 99DE-1032922.  
 PR 14-JUL-1999; 99DE-1032926.  
 PR 14-JUL-1999; 99DE-1032928.  
 PR 14-JUL-1999; 99DE-1033004.  
 PR 14-JUL-1999; 99DE-1033005.  
 PR 14-JUL-1999; 99DE-1033006.  
 PR 12-AUG-1999; 99US-0148613.  
 PR 27-AUG-1999; 99DE-1040764.  
 PR 27-AUG-1999; 99DE-1040765.  
 PR 27-AUG-1999; 99DE-1040766.  
 PR 27-AUG-1999; 99DE-1040832.  
 PR 31-AUG-1999; 99DE-1041378.  
 PR 31-AUG-1999; 99DE-1041379.  
 PR 31-AUG-1999; 99DE-1041380.  
 PR 31-AUG-1999; 99DE-1041394.  
 PR 31-AUG-1999; 99DE-1041396.  
 PR 03-SEP-1999; 99DE-1042076.  
 PR 03-SEP-1999; 99DE-1042077.  
 PR 03-SEP-1999; 99DE-1042079.  
 PR 03-SEP-1999; 99DE-1042086.  
 PR 03-SEP-1999; 99DE-1042087.

PR 03-SEP-1999; 99DE-1042088.  
 PR 03-SEP-1999; 99DE-1042095.  
 PR 03-SEP-1999; 99DE-1042124.  
 PR 03-SEP-1999; 99DE-1042129.  
 PR 09-MAR-2000; 2000US-0187970.  
 XX  
 PA (BADI ) BASF AG.  
 XX  
 XX Pompejus M, Kroeger B, Schroeder H, Zeider O, Habershauer G;  
 PI  
 DR WPI; 2001-137957/14.  
 DR N-PSDB; AAF71800.  
 XX  
 XX Nucleic acids from Corynebacterium glutamicum encoding metabolic  
 PT pathway proteins, useful for producing fine chemicals in  
 PT microorganisms, including organic acids, nonproteinogenic amino acids,  
 PT and purine and pyrimidine bases -  
 XX  
 PS Claim 20; Page 300-301; 1737pp; English.  
 XX  
 CC AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic  
 CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum  
 CC MP nucleic acids are useful for the production of fine chemicals  
 CC in microorganisms, including organic acids, nonproteinogenic amino  
 CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,  
 CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic  
 CC compounds, vitamins, cofactors, polyketides and enzymes.  
 XX  
 SQ Sequence 446 AA;  
 Query Match 19.8%; Score 470.5; DB 22; Length 446;  
 Best Local Similarity 28.9%; Pred. No. 6.1e-37;  
 Matches 134; Conservative 84; Mismatches 174; Indels 71; Gaps 12;  
 QY 15 GKIDTVVLACVDMQGRIMGKRLTGRHFLGLDQKQISISTVYAVTTEGIAGGVEISSVD 74  
 DB 29 GHLKSVVVAPEALESAL-----EEGIFD-----GSAIEGYA-----RISEAD 66  
 QY 75 TGYSDDCHLCADLNSLLPMSQG-AVLAINSPHFVT---SEPLFCSPRVILMQOIERLA 130  
 DB 67 T-----IARPPSTFQVLPLEAGISKLAARLFCVTPDQGSFSDPROVLRQVGLAA 121  
 QY 131 NLKLGFLFASLEFNLFNETYKSASQKHKNLKT-----AOPHHOMNISASSGIETFM 184  
 DB 122 DEGLTCWISPEIEFLV-----QSLRTNGLPPVPTDNGGYDQATFNEAPNFR 169  
 QY 185 RSVRNKLEBAGILMEATHPEFLPSQHELNPVADPLTMADRHIIAKHGVREMAEQSGVA 244  
 DB 170 RNAMVALEELGIEVFESHETAFQGOEIDLRHADALTMADNIMTFRYIMKQVARDQGVGA 229  
 QY 245 TFMAKLSSTALGNACHHMSLODAETKNAFYDQNDYGMSTLARNWIAGLLKVYPEATY 304  
 DB 230 SFMPKPFQEHAGSAMTHMSLFGDT--NAFHDPDSYMLSKTAKOFIAGILHAPFETA 287  
 QY 305 FFASYINSYKRLQPLTFAPTCCWAI DNRTSAFRLCN---SKSEGINVELIRIGGADLNPY 361  
 DB 288 VTNQWNSYKRIYVYNEAPTAATMGVSNRSALRVPTYLNKEESRRVRLPDTACNPY 347  
 QY 362 LAFSAIIAGISGIEBEKLELPPPASGNV-----YDKELPEFPNSIQNATHLL 409  
 DB 348 LAFSVMLGAGLKGIEGYELDEPAEDDISNLSFRERRAMGYND-----LPSSLDQALROM 402  
 QY 410 KESKMLNKTFGEKLIILHYVNAANVEINFEFSKVDTWELNQGPN 452  
 DB 403 EKSELVADILGEHVFEEFFLRNKWREWRDYOEQITPWEILRNLD 445  
 RESULT 7  
 ABB55555  
 ID ABB55555 standard; Protein; 446 AA.  
 XX  
 AC ABB55555;  
 XX



QY 65 GGGVEISSVD---TGYSDDCHLCADLNSHLPLPMSSEG-----AVLAISNPHNFVTSPL 114  
 D 114  
 D 48 GLGFDGSAIEGMTRVSEDDMIVQDPSTFQILPWRGGPGQTARMPCDILTP-----DGEPS 103  
 QY 115 FCSRVILMOQIERLANLKLGL---FASLEFNLFNETYKASQKHWNKLTAPQHQM 171  
 D 171  
 D 104 LGDPRHVLKR---TLAKAKDKGFTFYVHPTEIFLFE-----SQDDWSKAPTIDEQGY 154  
 QY 172 M-NISASGIEFTFMSRVNKLKEEAGILMEATHPEFLPSQHELNFVPADPLTMADRHILAK 230  
 D 230  
 D 155 FDHVPSPGMD-FRATVNMLEQWISVEYSHHAGPCQNEIDURYADALTMADNIMTFR 213  
 QY 231 HGVREMAQSGNVATFMAKLSSTALGNACHIMSLQDAETKNAFYQDNDYGMSTLARN 290  
 D 290  
 D 214 TVVKEISLERGIHASFPEKPLADAPGSGMTHLSL--PEGDSNAFYEAGQEFNMSLTARQ 271  
 QY 291 WIAGLLKVPYEAATFFASYINSYKRLQPLTAPYKCCWAIDNRTSAFRLCN---SKSEGI 347  
 D 347  
 D 272 FAAGILYHAAEICAVTDQVNSYKRLMGNEAPSYICWGHNNRSALLRIPQYKPKGNLSA 331  
 QY 348 NVELRIGGADLNPYLAFSAIIAAGISGIEEKLPLPPASGNVY--NDKE-----LPEPPN 400  
 D 400  
 D 332 RMEFRALDPVANPYLAYSVLLAAGLDGDKQWQCEPTSDDVWELTGERQAMGIQPLPR 391  
 QY 401 SLQNAHLKESKMLNTFGKELILHYVNAANVEINEFSPKQVTDWELNQ 449  
 D 449  
 D 392 SLDEALKIMKESDFADVGLGSHAFGYFLDNKHQEWEEYNQVTPVELKK 440

## RESULT 9

ABP27862

ID ABP27862 standard; Protein; 448 AA.

AC ABP27862;

XX 02-JUL-2002 (first entry)

DT Streptococcus polypeptide SEQ ID NO 4900.

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

XX group A streptococcus; Streptococcus pyogenes; antibacterial;

XX antiinflammatory; infection; vaccine; meningitis; gene therapy.

XX Streptococcus agalactiae.

XX WO200234771-A2.

XX 02-MAY-2002.

XX 29-OCT-2001; 2001WO-GB04789.

XX 27-OCT-2000; 2000GB-0026333.

XX 24-NOV-2000; 2000GB-0028727.

XX 07-MAR-2001; 2001GB-0005640.

XX (CHIR-) CHIRON SPA.

XX (GENO-) INST GENOMIC RES.

XX Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;

XX Tettelin H;

XX WPI; 2002-352536/38.

XX N-PSDB; ABN68493.

XX New Streptococcus protein for the treatment or prevention of infection

XX or disease caused by Streptococcus bacteria, such as meningitis, and

XX for detecting a compound that binds to the protein -

XX Claim 1; Page 3652; 4525pp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B

XX streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS

CC

CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in  
 CC the specification. The proteins have antibacterial and antiinflammatory  
 CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and  
 CC antibodies that bind (I) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
 CC biological sample. (I) is used to determine whether a compound binds to  
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (I) may be used to recombinantly produce (I) and may be  
 CC used in gene therapy. Antibodies to (I) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins.

XX Sequence 448 AA;

QY Query Match

D Best Local Similarity 19.7%; Score 445; DB 23; Length 448;

D Matches 131; Conservative 27.4%; Pred. No. 2e-34;

D Mismatches 80; Indels 54; Gaps 10;

QY 1 MTITYDELNNLRNGKIDTVVLCVDMQGRMLGKRLTGRHFLGLDQ---KKISISTFYA 57

D 57

QY 1 MTITAEIRREVKEKNVTFRLMFTDILGVNKNVEIP-----ATDEOLDKVLNKAHPDG 55

D 55

QY 58 VTIEGIAGGGYEISSVDYGYSDCHLCADLNSHLPLPMSSE--GAVLAISNPHNFVTSPLF 115

D 115

QY 56 SSIEGF-----VRINESDMVLYPDLOTWIVFPWGDENGAVAGLICDIYTAEGEPFA 106

D 106

QY 116 CSRVILMOQIERLANLKLGL-FASLEFNLFNETYKASQKHWNKLTAPQHQMNI 174

D 174

QY 107 GDRGNLKRNMKRMQEMGYKSNLGPPEPFLF-----KMDENGCPNPTLDV 151

D 151

QY 175 SASG-----IETFMRSVRNKLKEEAGILMEATHPEFLPSQHELNFVPADPLTMADR 225

D 225

QY 152 NDKGGYFDLAPTDLADNTREIIVNLTQMGEVASHHVAVGQHEIDFYDDVYLKACDN 211

D 211

QY 226 HIITAKHGVREMAEQSGMVATFMAKLSSTALGNACHIMSLQDAETKNAFYQDNDYGM- 284

D 284

QY 212 IQFLVVKVTKIARKGLYATFMAKPKPGINGSGHCHNMSLFDNE--GNAFFDPEDPRQM 270

D 270

QY 285 -STLARNIAGLLKYVPEATYFFASYINSYKRLQPLTAPYKCCWAIDNRTSAFRLCNSK 343

D 343

QY 271 LSEDAVYVFLGLMKHAYNTAIINPTVNSYKRLVPGVEAPVYVAVAGNRSPILRVPASR 330

D 330

QY 344 SEGIVNELRIGGADLNPYLAFSAIIAAGISGIEEKLPLPPASGNVY-----NDKELP 396

D 396

QY 331 GMGTRLELRSDVPTANPYLALSVLGSGLEGIEKIEAPEPIETNIYAMTVEERRQAGIV 390

D 390

QY 397 EFPNSLQNAHLKESKMLNKTGKELILHYVNAANVEINEFSPKQVTDWELNQGFNRY 454

D 454

QY 391 DLPSTLHNALEAEDEEVVKAALGTHIYTNFLQAKRIEWASYATYVSQWEIDNYLDLY 448

D 448

RESULT 10

ABP27863

ID ABP27863 standard; Protein; 448 AA.

XX ABP27863;

XX 02-JUL-2002 (first entry)

DT Streptococcus polypeptide SEQ ID NO 4902.

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

XX group A streptococcus; Streptococcus pyogenes; antibacterial;

XX antiinflammatory; infection; vaccine; meningitis; gene therapy.

XX Streptococcus pyogenes.

XX WO200234771-A2.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX



Matches 128; Conservative 83; Mismatches 220; Indels 40; Gaps 10;

QY 1 MTITYDELNNLRNGKIDTVVLACVDMQGRMLGKRLTGRHFLGLDQ---KKISISTFVYA 57  
 DB 1 MPITAADIRREVKEKNVTEIRLFMSDILGTMKVNEIP-----ATDEQLDKVLNKNVDFG 55

QY 58 VTIEGIAGGGEISSVDGTSDCHLCADLNSHLPLPWE--GAVLAISNPHNFVTSEPLF 115  
 DB 56 SSIEGF-----VRINESDMYLPDLDTWTVPFGDENGVSAGLICDVYTTGEPPA 106

QY 116 CSPRVILMQOIERLANLKLKGL--FASELEFNLF--NETYKSASQKHWNKLTQAOPHHQWM 172  
 DB 107 GDRPNLKRALRHMEEVGFKNLGPPEFFFLKLDENGDPTEVNDKG-----GYF 158

QY 173 NISASSGIETFMRSVRNKLBEAGILMEATHPEFLPSQHELNFPADPLTMADRHIIAKHG 232  
 DB 159 DLAPTDLADNTRREIVNVLTKMGFEVASHHEVAVGQHEIDFKYDEVLRACDKIQLFKLV 218

QY 233 VREMAEQSGMVATFMAKLSSTALGNACHIHMSLODAETEKNAFYDQNDGYM--STLARN 290  
 DB 219 VKTIARKHGLYATFMAKPKFGIAGSGMHCNMSLFDAD--GNNAFFDNDPKGMQLSETAYH 277

QY 291 WIAGLLKYVPEATYFFASYNSYKRLQPLTFAPTKCCWAIDNRTSAFRLCNSKSEGINVE 350  
 DB 278 FLGLLKHAYNYTAIMNPTVNSYKRLVPGYEAPVYIAWAGNRSPLVVRVPASRGMGTRLE 337

QY 351 LRIGGADLNPYLAFAISAIAGISGIEEKLPLPPASGNVY-----NDKELPEFPNSLQ 403  
 DB 338 LRSVDPMPANPYVAMAVLLEVLGYLGIENKIEAPAPIENIYIMTAERKEAGITDLPSTLH 397

QY 404 NATHLLKESKMLNKTGFEKLIHLYVNAANVEINERFSQVTDWELNQGFNRY 454  
 DB 398 NALKALTEDEWKAALGDHIYTSFLEAKRIEWASATFVSQWEIDNYLDLY 448

## RESULT 12

AAE33506  
 ID AAE33506 standard; Protein: 448 AA.

XX AC AAE33506;

XX DT 02-APR-2003 (first entry)

XX DE Streptococcus pneumoniae ID-216 protein.

XX KW Vaccine; therapy: Streptococcus pneumoniae infection; gene therapy;  
 antibacterial; immunostimulant.

XX OS Streptococcus pneumoniae.

XX FH Key Location/Qualifiers

XX FT Misc-difference 97

XX FT /note= "Encoded by GTTC"

XX PN W0200279241-A2.

XX PD 10-OCT-2002.

XX PF 28-MAR-2002; 2002W0-GB01480.

XX PR 30-MAR-2001; 2001GB-0008079.

XX PA (MICR-) MICROBIAL TECHNIKS LTD.

XX PA (PROV-) PROVALIS UK LTD.

XX PI Le Page RWF, Badcock D, Sizer PJH, Peek K, Wells JM, Hanniffy SB;

XX DR WPI; 2003-103261/09.

XX DR N-PSDB; AAD51270.

XX PT New Streptococcus pneumoniae protein or polypeptide, useful as an  
 immunogen and/or antigen for use in vaccines against Streptococcus  
 pneumoniae infection, and in diagnostic assays

XX Claim. 7; Page 38; 43pp; English.

XX The invention relates to Streptococcus pneumoniae protein or polypeptide  
 and their corresponding nucleic acid sequences. The invention also  
 relates to their use in vaccines and in screening methods. The invention  
 is useful as an immunogen and/or antigen that may be used in vaccines or  
 diagnostic assays. The methods are useful for the selection/diagnosis of  
 S. pneumoniae, and determining whether a protein or polypeptide  
 represents a potential anti-microbial target. An agent capable of  
 antagonising, inhibiting or otherwise interfering with the function or  
 expression of a protein or polypeptide is useful in the manufacture of  
 a complement for use in the treatment or prophylaxis of S.pneumoniae  
 infection. The invention is also used in gene therapy. The present  
 sequence is S. pneumoniae protein of the invention.

XX Sequence 448 AA;

Query Match 18.2%; Score 432; DB 24; Length 448;  
 Best Local Similarity 27.2%; Pred. No. 3.7e-33;  
 Matches 128; Conservative 83; Mismatches 220; Indels 40; Gaps 10;

QY 1 MTITYDELNNLRNGKIDTVVLACVDMQGRMLGKRLTGRHFLGLDQ---KKISISTFVYA 57

DB 1 MPITAADIRREVKEKNVTEIRLFMSDILGTMKVNEIP-----ATDEQLDKVLNKNVDFG 55

QY 58 VTIEGIAGGGEISSVDGTSDCHLCADLNSHLPLPWE--GAVLAISNPHNFVTSEPLF 115

DB 56 SSIEGF-----VRINESDMYLPDLDTWTVPFGDENGVSAGLICDVYTTGEPPA 106

QY 116 CSPRVILMQOIERLANLKLKGL--FASELEFNLF--NETYKSASQKHWNKLTQAOPHHQWM 172

DB 107 GDRPNLKRALRHMEEVGFKNLGPPEFFFLKLDENGDPTEVNDKG-----GYF 158

QY 173 NISASSGIETFMRSVRNKLBEAGILMEATHPEFLPSQHELNFPADPLTMADRHIIAKHG 232

DB 159 DLAPTDLADNTRREIVNVLTKMGFEVASHHEVAVGQHEIDFKYDEVLRACDKIQLFKLV 218

QY 233 VREMAEQSGMVATFMAKLSSTALGNACHIHMSLODAETEKNAFYDQNDGYM--STLARN 290

DB 219 VKTIARKHGLYATFMAKPKFGIAGSGMHCNMSLFDAD--GNNAFFDNDPKGMQLSETAYH 277

QY 291 WIAGLLKYVPEATYFFASYNSYKRLQPLTFAPTKCCWAIDNRTSAFRLCNSKSEGINVE 350

DB 278 FLGLLKHAYNYTAIMNPTVNSYKRLVPGYEAPVYIAWAGNRSPLVVRVPASRGMGTRLE 337

QY 351 LRIGGADLNPYLAFAISAIAGISGIEEKLPLPPASGNVY-----NDKELPEFPNSLQ 403

DB 338 LRSVDPMPANPYVAMAVLLEVLGYLGIENKIEAPAPIENIYIMTAERKEAGITDLPSTLH 397

QY 404 NATHLLKESKMLNKTGFEKLIHLYVNAANVEINERFSQVTDWELNQGFNRY 454

DB 398 NALKALTEDEWKAALGDHIYTSFLEAKRIEWASATFVSQWEIDNYLDLY 448

## RESULT 13

ABU00865

ID ABU00865 standard; Protein: 448 AA.

XX AC ABU00865;

XX DT 11-FEB-2003 (first entry)

XX DE S. pneumoniae type 4 strain protein from coding region #433.

XX KW Bacterial meningitis; pneumonia; sepsis; otitis media;  
 ear infection; anti-inflammatory; antibacterial; immunostimulant;  
 auditory; respiratory; gene therapy; vaccine.

XX OS Streptococcus pneumoniae type 4 strain.

XX PN W0200277021-A2.

XX



CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence represents an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC <http://wipo.int/pub/published/pct/sequences>.

[illegible]

RESULT 15	
AAU36745	
ID	AAU36745 standard; Protein; 446 AA.
XX	
XX	AAU36745;
XX	
XX	14-FEB-2002 (first entry)
XX	
XX	Staphylococcus aureus cellular proliferation protein #915.
DE	
XX	
XX	Antisense; prokaryotic cellular proliferation protein;
KW	antibiotic; antibacterial; drug design.
KW	
XX	
XX	Staphylococcus aureus.
OS	
XX	
XX	WO200170955-A2.
PN	
XX	
XX	27-SEP-2001.
PD	

PR	22-DEC-2000; 2000US-257931P.
XX	16-FEB-2001; 2001US-269308P.
XX	
XX	(ELIT-) ELITRA PHARM INC.
XX	
PI	Haseibeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI	Yamamoto RT, Xu HH;
PI	WPI; 2001-6111495/70.
DR	N-PSDB; AAS54604.
DR	
XX	
XX	New polynucleotides for the identification and development of
PT	antibiotics, comprise sequences of antisense nucleic acids -
XX	
XX	Example 3; Seq ID No 12338; 511pp; English.
PS	
XX	
XX	The invention relates to antisense inhibitors of genes essential to
CC	prokaryotic cellular proliferation, their use in identifying the
CC	genes, their use in the discovery of novel antibiotics, the essential
CC	genes themselves and the encoded proteins. The prokaryotes used are
CC	Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC	pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC	invention is also useful for the identification of potential new targets
CC	for antibiotic development. The antisense nucleic acids can also be used
CC	to identify proteins used in proliferation, to express these proteins,
CC	and to obtain antibodies capable of binding to the expressed proteins.
CC	The proteins can be used to screen compounds in rational drug discovery
CC	programmes. The antisense nucleic acid sequence is also useful to screen
CC	for homologous nucleic acids which are required for cell proliferation in
CC	a wide variety of organisms. The present sequence represents an
CC	essential prokaryotic cellular proliferation protein.
CC	Note: The sequence data for this patent did not form part
CC	of the printed specification, but was obtained in electronic
CC	format directly from WIPO at
CC	ftp.wipo.int/pub/published pct sequences.

Query Match	17.9%	Score	424.5;	DB	22;	Length	446;
Best Local Similarity	26.2%;	Pred.	No. 2e-32;				
Matches	123;	Conservative	92;	Mismatches	211;	Indels	43;
Gaps	9						

  

QY	2	TITYDELNNLRNGKIDTVVCLAVCDMGRLMKRKLTCGRHFLGLDQKKISITSTFFVAVATIE	61
DB		: : : :     : : : :     : : : :	
DB	5	TFTKDDIRKPAEENRYLRLQTDLGTTIKNVEVPVSQL----EKVLDNEMMFDSGISIE	60
QY	62	GIAGGGYEISSVDGTGYSDCHLCADLNLSHLPLWS--EGAVLAISNPHNFTSEPLFCSPR	119
DB		: : : :     : : : :     : : : :	
DB	61	GF-----VRIESDMYLHPDLDTWTWTFPMTAGQGKVARLICDVYTKDGTPEFGDPR	111
QY	120	VILMOOIERLANLKLAGL-FASELEFNLFNETKYSASOKHWKNLKTAAQHQQ-----WM	172
DB		: : : :     : : : :     : : : :	
DB	112	ANLKRVLKEMEDLGFTDFNLGPEPEFPLFKLDEK-----GEPTLELNDGCGYF	159
QY	173	NISASSGITFTFMRSVRNKLEEAGILMEATHPEFLPSQHELNFVPADPLTWADRHHIAKHG	232
DB		: : : :     : : : :     : : : :	
DB	160	DLAPTDLGCNRDRDIVLEIEDMCFDIASHHEVAPGQHEIDFYKADAVTACDNIQTPKLV	219
QY	233	VREMAEQSGMVATFMNAKLSTALGNACHIHMSLDAAETEKNAVFDONDEVGMSTLARWI	292
DB		: : : :     : : : :     : : : :	
DB	220	VKTIARKHNLHAFTMPKPFLFVGNSGWHFNVSU--FKGKENAFDPDPNTMGLTETAVOFT	277
QY	293	AGLLKYVPEATYFAFSYINSYKRLQPLTPAPTKCCWAI DNRTSAFRCLNSKSEGIVNELR	352
DB		: :         : :     : : : :	
DB	278	AGVLKNARGFTVCNPLVNSYKRLVGYEAPCVIANGSKQRSPLIRVPSSRGSLSTRIBVR	337
QY	353	IGGADLNPLYAFSAIIAAGSITSEEKLELPPLPPASGVNY--NDKE-----LPEFPNSLQNA	405
DB		: : : :     : : : :     : : : :	
DB	338	SVDPAANPYMALAAIIIEAGLDGINKVKVPEPVNQNIYENRREAREAVGIQDLFSTLYTA	397
QY	406	THLLKESKMNLKTFGBKLI LHYVNAANVEINFEFSQVTDMLNKGFNRY	454
DB		: : : :     : : : :     : : : :	
DB	398	LKAMRENVIKKALGNHIYNQFTNSKSIEDWDYTRTVQSVBERDOYMQRQY	446

Search completed: December 17, 2003, 22:42:57  
Job time : 43 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 17, 2003, 22:51:30 ; Search time 5280 Seconds  
(without alignments)  
3517.608 Million cell updates/sec

Title: US-10-098-602A-2  
Perfect score: 2376  
Sequence: 1 MTITYDELNLNRKIDTV.....INEFSKQVTDWELNQGFNRY 454

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US10098602/runat\_17122003\_151236\_28042/app\_query.fasta\_1.647  
-DB=GenEmbl -QWTF=fastcap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=500 -THR MAX=100 -ALIGN=15 -MODE=LOCAL  
-OUTPMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10098602@cgn\_1.1 3508 @runat\_17122003\_151236\_28042 -NCPU=6 -ICPU=3  
-NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:  
1: gb.ba:\*  
2: gb.htg:\*  
3: gb.in:\*  
4: gb.om:\*  
5: gb.ov:\*  
6: gb.pat:\*  
7: gb.ph:\*  
8: gb.pl:\*  
9: gb.pr:\*  
10: gb.ro:\*  
11: gb.sts:\*  
12: gb.sy:\*  
13: gb.un:\*  
14: gb.vi:\*  
15: em.ba:\*  
16: em.fun:\*  
17: em.hum:\*  
18: em.in:\*  
19: em.mu:\*  
20: em.or:\*  
21: em.ov:\*  
22: em.pat:\*  
23: em.ph:\*  
24: em.pl:\*  
25: em.ro:\*  
26: em.sts:\*  
27: em.un:\*  
28: em.un:\*

29: em.vi:\*  
30: em.htg.hum:\*  
31: em.htg.inv:\*  
32: em.htg.other:\*  
33: em.htg.mus:\*  
34: em.htg.pln:\*  
35: em.htg.rod:\*  
36: em.htg.mam:\*  
37: em.htg.vrt:\*  
38: em.sy:\*  
39: em.htgo.hum:\*  
40: em.htgo.mus:\*  
41: em.htgo.other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	1015	42.7	345783	1	AP003001 Mesorhizo
C 2	969.5	40.8	11725	1	AE014569 Brucella
C 3	963	40.5	323450	1	SMES91790 Sinochizo
C 4	949.5	40.0	12340	1	AE009691 Brucella
C 5	942.5	39.7	10069	1	AE009164 Agrobacte
C 6	942.5	39.7	11253	1	AE008130 Agrobacte
C 7	799.5	33.6	13246	1	MTV003
C 8	799.5	33.6	14715	1	AE007117 Mycobacte
C 9	799.5	33.6	318050	1	BX248344 Mycobacte
C 10	798.5	33.6	277000	1	SC0939109 Streptomy
C 11	773.5	32.6	298550	1	AP005047 Streptomy
C 12	614.5	25.9	1860	6	AX259711 Sequence
C 13	614.5	25.9	11355	1	PSP311159 Pseudomon
C 14	595	25.0	128136	1	AF440524 Pseudomon
C 15	554	23.3	300550	1	AP005021 Streptomy
C 16	510.5	21.5	305153	1	AP001520 Bacillus
C 17	510	21.5	310266	1	AE016862 Pseudomon
C 18	509	21.4	300242	1	AE016790 Pseudomon
C 19	493.5	20.8	1338	6	BD180047 Highly th
C 20	487	20.5	10453	1	AE004758 Pseudomon
C 21	481	20.2	18194	1	AE001758 Thermotog
C 22	477	20.1	1444	1	TMGLNA
C 23	475.5	20.0	5500	6	AX503498 Sequence
C 24	475.5	20.0	5500	6	BD177689 Process f
C 25	470.5	19.8	1338	6	AX122526 Sequence
C 26	470.5	19.8	1338	6	BD164643 Novel pol
C 27	470.5	19.8	1461	6	AX063813 Sequence
C 28	470.5	19.8	6433	1	CGL310086 Corynebac
C 29	470.5	19.8	306650	1	AP005221 Corynebac
C 30	470.5	19.8	326150	1	AP005281 Corynebac
C 31	470.5	19.8	349980	6	AX127150 Sequence
C 32	466	19.6	2192	1	CLOGLSA
C 33	466	19.6	142665	1	AE016794 Pseudomon
C 34	465	19.6	10458	1	AE009688 Brucella
C 35	465	19.6	13418	1	AE014571 Brucella
C 36	465	19.6	310613	1	AE016875 Pseudomon
C 37	464.5	19.5	11906	1	AE001904 Deinococc
C 38	464	19.5	12185	1	AE006449 Lactococc
C 39	463	19.5	43883	2	AC004396 Pseudomon
C 40	459	19.3	12795	1	AE004467 Pseudomon
C 41	458	19.3	9960	1	AE004630 Pseudomon
C 42	453	19.1	10876	1	AE014758 Bifidobac
C 43	453	19.1	11364	1	AE012342 Xanthomon
C 44	453	19.1	305245	1	AE016793 Pseudomon
C 45	453	19.1	349980	6	AX492786 Sequence

## ALIGNMENTS

RESULT 1

```

AP003001/c
LOCUS      345783 bp      DNA      linear      BCT 15-MAY-2001
DEFINITION Mesorhizobium loti DNA, complete genome, section 8/21.
ACCESSION AP003001 BA000012.
VERSION    AP003001.2 GI:14023393
KEYWORDS
SOURCE
ORGANISM   Mesorhizobium loti
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Phyllobacteriaceae; Mesorhizobium.
REFERENCE
AUTHORS    Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S.,
            Watanabe, A., Iidesawa, K., Ishikawa, A., Kawashima, K., Kimura, T.,
            Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuono, A.,
            Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimpou, S., Sugimoto, M.,
            Takeuchi, C., Yamada, M. and Tabata, S.
            Complete genome structure of the nitrogen-fixing symbiotic
            bacterium Mesorhizobium loti
            JOURNAL   DNA Res. 7 (6), 331-338 (2000)
MEDLINE    21082930
PUBMED     11214968
REFERENCE
AUTHORS    Kaneko, T.
            Direct Submission
            Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
            Institute, The First Laboratory for Plant Gene Research; Yana
            1532-3, Kisarazu, Chiba 292-0812, Japan
            (E-mail: kaneko@kazusa.or.jp,
            URL: http://www.kazusa.or.jp/rhizobase/,
            Tel: 81-438-52-3935 (ex. 2338), Fax: 81-438-52-3934)
            On May 11, 2001 this sequence version replaced gi:11994976.
FEATURES
source     Location/Qualifiers
            1..345783
             /organism="Mesorhizobium loti"
             /mol_type="genomic DNA"
             /strain="MAFF303099"
             /db_xref="taxon:381"
             complement(6..884)
             /gene="mlr3011"
             complement(6..884)
             /gene="mlr3011"
             /codon_start=1
             /transl_table=11
             /product="transcription regulator"
             /protein_id="BAB50000.1"
             /db_xref="GI:14023394"
             /translation="MDRLMAMSTFIRVIETGFSAAARHLRVQGSNAVAQLERL
             GRLLSRTTRGSLPTDAGLSFVEGAKRAVDDEADEAENAAAGTAALTLGLRIGTSVTF
             GRLLVPRKAGLDHPRLDVEVAMDRNVDLVEGVDMRLRGLSEFAPAHIGRKIAA
             CRLVVAAPSLAREGEPTAPDLSRHEAILFVRAGTSIRWTFRGAEVSLAGQRL
             RYTSAGIRENLVSGIGFTVGSEWFDALRSRGAVRPILTEWLPVELWALFPAGRK
             ASARATVEFLAGSL"
             994..1740
             /gene="mlr3013"
             994..1740
             /gene="mlr3013"
             /codon_start=1
             /transl_table=11
             /product="short chain dehydrogenase or 3-ketoacyl-CoA
             reductase"
             /protein_id="BAB50001.1"
             /db_xref="GI:14023395"
             /translation="MGKLNKNAIVTGAARGIGAGTAKLAELAEAGAANNVYNATSRRAA
             ERTKSVIAGGKATAVQGVDAEAEVEKFAAEIAPGRKSLINNAATASFGTFEE
             ASGEAYRRLFDNLVTLTILTKAAMRHFPETGSIIVNIGTISWNPNTSYSAKKA
             AIDLTLISLRGARNIRVNIAPVQYTHTEMTGEGVGTDFGNMLIAGVPLQGRFGKP
             DDIAPTVAFLADEAAWLGTGERINAGGGR"
             1914..3371
             /gene="mlr3014"
             1914..3371
             /gene="mlr3014"
             /codon_start=1
             /transl_table=11

/product="homospersmidine synthase"
/protein_id="BAB50002.1"
/db_xref="GI:14023396"
/translation="MLDKMANENHPVYGEITGPVVMIGFSGSIGCTLPLIERHFKFDK
SRMTVLDPDRDTRKLLDGRGAFVQEAVENTENNYKLLPLLTNGGGQGCFCVNLSDVTG
SVDMLRCRLKGLVYIDTVBFWLGFYFDADAKNADNRNTYALREAMIKEKDKPGGAT
VVTGCGANPMGVSWFKQALVNLATDLGLEFSEPAQEDREGAKMLKMGKVGKGIHAE
RDORTKKPKMDVFWNTWSVEGFTISEGLQPAELGWGTHEKMPKNGKGGKHGKSAAI
YLEOPGANTVRSMCTPGACQGLLVTHNEAISIAIDFTVRSKKGKGVVRPTCHVAYH
PSNDAMLSLDEMFGAAGKAPQVHHVLDENELVDGDELGLLYLGHDKNAYWGSLSL
AEARKLADYQNAQTGMQVTSVLAAGVMVALENPDAGIVEADEMDYKRCLDVQSPYLGVP
KGYTDMTPFLDRRPLGFPEDLDRSDPQWQFRNILVR"
3431..3931
/gene="mlr3016"
3431..3931
/gene="mlr3016"
/codon_start=1
/transl_table=11
/product="outer membrane lipoprotein"
/protein_id="BAB50003.1"
/db_xref="GI:14023397"
/translation="MRPGAACVLDRDARPFQHFGRSHRACQNRNRLDRCVWRGFMQFTI
ARKVLSIGMAGSFAMLAACVTSQPDAPPMAAPKGVGSGWIDAKGTGLSTPTAGKFA
TVATDTGQKLADGSYTMGTGATSVETNGTSLIRQTPVFNCLLISTSQLNCTSSSQGNF
VLTRRT"
4123..6243
/gene="mlr3017"
4123..6243
/gene="mlr3017"
/EC_number="3.1.3.5"
/codon_start=1
/transl_table=11
/product="s'-nucleotidase"
/protein_id="BAB50004.1"
/db_xref="GI:14023398"
/translation="MKKLAALAAALSVTLGLSAGASFADYTLNLIHFNMDHWSIRGN
KYESTCSADEETKGEICGAGRLITAIQAERKKLEGQNVLLNAGDSFGSLFYTTYK
GTVEEFLNOMKPDVTLGNHEFDGSEALVPYLDKAKFPVLSANMPPDKSAAAGKI
KPSIVVEVGOKIGIVGAVTNDTPELASPGNIIATDDVESTATVEKLKAQGVNKII
AVTHIGYNRRDIIAKIPGVVVVGGHSHLLSNTDPAAGPYPTMNDPGYKVPVV
QAASYSKYLGEFVVDNDGVVKSASGDFIUDKSIITDPPAVLARIKELGALIEALKN
KEVAETTDVIDGSRNCRACEMGNLVDALIDRVKQGVIEIVISNGGLRASIDKG
VTVMGEVLTPFQNTLATFQISGKDLVAGLESLSQIEDGAGRFQVAGLKYSDKS
VAPNAGRVKSVSEVMENGAWTPIKPKDYLVTATNVYVROGGDGYKYPFAERAKNAYDGP
GLEOVVADYLGARHPYTPKLDGRITETIAATVAAAAPAEPAKPAEATPAPAPAKPA
EAKPAEPAKPAEAPAEPAKPAEPAKPAEPAKPAEPAKPAEPAKPAEPAKPAEPAK
EAPAPAEPAKPAEPAKPAEPAKPAEPAKPAEPAKPAEPAKPAEPAKPAEPAKPAE
IYEANKGQRPRLTTPGATLTTPAK"
6381..7442
/gene="mlr3019"
6381..7442
/gene="mlr3019"
/EC_number="4.99.1.1"
/codon_start=1
/transl_table=11
/product="ferrochelatase"
/protein_id="BAB50005.1"
/db_xref="GI:14023399"
/translation="MTSAIADVPAKAAKPVGTAGHAPAGAGKGVLLVNLGTPDGTEF
KPMRYRLREFLSDPRVIELNKAIWYPILYGLVLTTPRKSGANYARIWNQERNESPLR
VTYRAQSEKLAEALDGLDPTVTDWAMRYGNPSTASVARELVAQGCDRILSFPLYQYS
ITTSANGIDQLFRALMKLRRAPIRSVPVYAEVPIEALASSIERHLATLDFEPEV
ATTVHGIPKPYDKGDPYQACHLETLLREKLGWDEKLLITTFOSRGAQEMLOPYT
DKTVSKLQKDGKYSIAIYNPFGSVDCITLDEIGREAAETFHAGSKGPAHIFCLNDS
ABGMVTIEMVVRREUSGWV"
7615..8826
/gene="mlr3020"
7615..8826
/gene="mlr3020"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB50006.1"

```



QY 404 AsnAlaThrHisLeuLeuLysGluSerLysMetLeuAsnLysThrPheGlyGluLysLeu 423  
 Db 60783 GAGCCACCGAAGCTGCGCAAGTCTGAAGATGCTGAACAGCGCTCGCGGAGGATGG 60724  
 QY 424 IleLeuHisTyrValAsnAlaAsnValGluIleAsnGluPheSerLysGlnValThr 443  
 Db 60723 CTCGAACATTATGTCACACCGCAAGTGGAGCAGTTCGAATACGACCGCGCATACG 60664  
 QY 444 AspTyrGluLeuAsnGlnGlyPheAsnArgTyr 454  
 Db 60663 GATTGGGAATCTGCACAGGGGGTTCGAGCGATAC 60631

## RESULT 2

LOCUS AE014569 11725 bp DNA linear BCT 02-OCT-2002  
 DEFINITION Brucella suis 1330 chromosome II section 65 of 108 of the complete genome.  
 ACCESSION AE014569  
 VERSION AE014569.1  
 KEYWORDS GI:23464092  
 SOURCE Brucella suis 1330  
 ORGANISM Brucellaceae; Brucella.

REFERENCE 1 (bases 1 to 11725)  
 AUTHORS Paulsen, I., Seehadri, R., Nelson, K.E., Eisen, J.A., Heidelberg, J.F., Read, T.D., Dodson, R.J., Umayam, L.A., Brinkac, L.M., Beanan, M.J., Daugherty, S.C., Deboy, R., Durkin, A.S., Kolonay, J.F., Madupu, R., Nelson, W.C., Ayodeji, B., Kraul, M., Shetty, J., Malek, J.A., Van Aken, S.B., Riedmiller, S., Tettelin, H., Gill, S., White, O., Salzberg, S.L., Hoover, D.L., Lindler, L., Halling, S.M., Boyle, S.M. and Fraser, C.M.  
 TITLE The Brucella suis genome reveals fundamental similarities between animal and plant pathogens and symbionts  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (20), 13148-13153 (2002)  
 PUBMED 12271122  
 REFERENCE 2 (bases 1 to 11725)  
 AUTHORS Paulsen, I., Seehadri, R., Nelson, K.E., Eisen, J.A., Heidelberg, J.F., Read, T.D., Dodson, R.J., Umayam, L.A., Brinkac, L.M., Beanan, M.J., Daugherty, S.C., Deboy, R., Durkin, A.S., Kolonay, J.F., Madupu, R., Nelson, W.C., Ayodeji, B., Kraul, M., Shetty, J., Malek, J.A., Van Aken, S.B., Riedmiller, S., Tettelin, H., Gill, S., White, O., Salzberg, S.L., Hoover, D.L., Lindler, L., Halling, S.M., Boyle, S.M. and Fraser, C.M.  
 TITLE Direct Submission  
 JOURNAL Submitted (14-AUG-2002) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA

FEATURES  
 source  
 1. .11725  
 /organism="Brucella suis 1330"  
 /mol\_type="genomic DNA"  
 /strain="1330"  
 /db\_xref="taxon:204722"  
 /chromosomes="II"  
 406. .564  
 /locus tag="BRA0728"  
 406. .564  
 /locus tag="BRA0728"  
 /note="similar to GP:15074478; identified by sequence similarity; putative"  
 /codon start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AA033910.1"  
 /db\_xref="GI:23464093"  
 /translation="MVRKNLIRSYNNWRRYRNTVNELSRLSPREINDLILPSEIIP VARKAAR"  
 complement (746. .1975)  
 /locus\_tag="BRA0729"  
 complement (746. .1975)  
 /locus\_tag="BRA0729"  
 /note="identified by match to TIGR protein family HMM TIGR01424"

gene  
 CDS  
 complement (5280. .5396)  
 /locus\_tag="BRA0733"  
 complement (5280. .5396)  
 /locus\_tag="BRA0733"  
 /note="identified by Glimmer2; putative"  
 /codon start=1  
 /transl\_table=11

gene  
 CDS  
 complement (5280. .5396)  
 /locus\_tag="BRA0733"  
 complement (5280. .5396)  
 /locus\_tag="BRA0733"  
 /note="identified by Glimmer2; putative"  
 /codon start=1  
 /transl\_table=11

gene  
 CDS  
 complement (5280. .5396)  
 /locus\_tag="BRA0733"  
 complement (5280. .5396)  
 /locus\_tag="BRA0733"  
 /note="identified by Glimmer2; putative"  
 /codon start=1  
 /transl\_table=11

gene  
 CDS  
 complement (5280. .5396)  
 /locus\_tag="BRA0733"  
 complement (5280. .5396)  
 /locus\_tag="BRA0733"  
 /note="identified by Glimmer2; putative"  
 /codon start=1  
 /transl\_table=11

gene  
 CDS  
 complement (5280. .5396)  
 /locus\_tag="BRA0733"  
 complement (5280. .5396)  
 /locus\_tag="BRA0733"  
 /note="identified by Glimmer2; putative"  
 /codon start=1  
 /transl\_table=11

gene  
 CDS  
 complement (5280. .5396)  
 /locus\_tag="BRA0733"  
 complement (5280. .5396)  
 /locus\_tag="BRA0733"  
 /note="identified by Glimmer2; putative"  
 /codon start=1  
 /transl\_table=11

gene  
 CDS  
 complement (5280. .5396)  
 /locus\_tag="BRA0733"  
 complement (5280. .5396)  
 /locus\_tag="BRA0733"  
 /note="identified by Glimmer2; putative"  
 /codon start=1  
 /transl\_table=11

gene  
 CDS  
 complement (5280. .5396)  
 /locus\_tag="BRA0733"  
 complement (5280. .5396)  
 /locus\_tag="BRA0733"  
 /note="identified by Glimmer2; putative"  
 /codon start=1  
 /transl\_table=11

gene  
 CDS  
 complement (5280. .5396)  
 /locus\_tag="BRA0733"  
 complement (5280. .5396)  
 /locus\_tag="BRA0733"  
 /note="identified by Glimmer2; putative"  
 /codon start=1  
 /transl\_table=11

/codon start=1  
 /transl\_table=11  
 /product="pyridine nucleotide-disulphide oxidoreductase family protein"  
 /protein\_id="AA033911.1"  
 /db\_xref="GI:23464094"  
 /translation="MDDRCIIIGAGHAGSQAASVLRQEGYAGEIILINDETDIPYHKP PLSKYLKAPKSGSLVRPESAYGNNNIEMLFGAHVDVASTAERTVTIGDGRVLPMSE LVPATGARAPIPDVPGVALEGVTLRMEADARRIAAMPDVNVVVIIGGGFGLGEMAH SAIALGKKTULIEAAPRVIGRSVAPSAHVAKRSRAADITLLTGLGLIASLEGNGRA IGVTAGDGTFFPADLVIGTGA VNPVELAAKAGLSIDNGIRVDEHMRTSAPHVYAIGD CASYAHFHAGHVRLESQNAQDQAKHVARTIVGRETPREVAMFWSQDGMKLTQTAG LSPDADRHLVSLSEPEENAFSVFHFGRDLRAVDSDINRPADHMIARRLLAAGTSPSEAD IMAGMPRLKEILLAVVAR"  
 complement (2087. .3031)  
 /locus\_tag="BRA0730"  
 complement (2087. .3031)  
 /locus\_tag="BRA0730"  
 /note="similar to GP:15024826; identified by sequence similarity; putative"  
 /codon start=1  
 /transl\_table=11  
 /product="transcriptional regulator, putative"  
 /protein\_id="AA033912.1"  
 /db\_xref="GI:23464095"  
 /translation="WLRFFHLCAIADYLIRWIFCOTQFAONKWAAGMETAAPMSAS VAEINERIDAMPAGERSAAQMLIANYPMLGLKTVAESESTAGVSSPTILRFVRLQGF QSYDPFQARLODELAQAQSPHRRNEQDKRPPDAADPGVEHTLENIRETFHRLGKQL HQVVASISLAKNGVLYIGGRTDPDIAQYTHMSIVCARVFLHGGQESMWRDLDDMG KRLDVLVDIRRYQESLLAAEKARSQGVVLVTQDMLSPIARVHVLGRIAPVS AWSAALFVLAETIIREMTRMLEKSDAARIAEALEALR"  
 3058. .3855  
 /locus\_tag="BRA0731"  
 3058. .3855  
 /locus\_tag="BRA0731"  
 /note="identified by Glimmer2; putative"  
 /codon start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AA033913.1"  
 /db\_xref="GI:23464096"  
 /translation="MANERDRAACEPVFVINEGESPYLLVCEHSHNHPESPAGL ELDATRTKTHAWDPGAVEARELSKALDAPLVEAGLSRLIDCNRPDADPLIPEIS GTTAVPGNHGLNTARMAVIDLSHRPFRHRIIEEIVAMEAARGQSWIVTVHSFTPVYR GVSPPWQIGIHDDEDDRIARPLIAALQDMLGHIGVNEPSPADRVVYTLERHARPRN APCVNIENDEITDAGTQTAWGRKLTALAKYIOETMRSAGAIPYA"  
 3919. .5286  
 /locus\_tag="BRA0732"  
 3919. .5286  
 /locus\_tag="BRA0732"  
 /note="identified by match to PFAM protein family HMM PF00120"  
 /codon start=1  
 /transl\_table=11  
 /product="glutamine synthetase family protein"  
 /protein\_id="AA033914.1"  
 /db\_xref="GI:23464097"  
 /translation="MAGQDTFDELKKAABRIDTVLACIVDMQRLIGKRPYGGFFV ESGYDETHGCNLLADIDMEFPVGYEAGWDKGYDFVMPKPDLSLRLAPLWLEKTAI VLDVLDHDDLSHSRAVLKQVQLRHERGYRAYFASELEFYIFDETYKSARAKRW HEMETASPYQDYYVTLTTRPEPVLRAMRNHLADAGIPVENSKGEMGQOQLANVRC KALEMADPHVIMKNAMKEIAEHGKCLITFMAYDYVARAGSSSHVHNSIWSADGKPLF FDPKAPYMTMLRMSWAGQIKYATDITYFLAPYINSYKRFQAGTTPATKIKWMSQDNR DAFRFLCGEGTGKIRIECRIGGIDINPYLAALIAAGLKGVDEKJDEDFEPVGDAYS AVRLEKIPYTLREAAALKGSALKEAFGEDVNVNHYTHATHEWQIEYDRRVTDWELRY GFRY"  
 complement (5280. .5396)  
 /locus\_tag="BRA0733"  
 complement (5280. .5396)  
 /locus\_tag="BRA0733"  
 /note="identified by Glimmer2; putative"  
 /codon start=1  
 /transl\_table=11



```
QY 341 AsnSerLysSerGluGlyIleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsnPro 360
Db 4942 GCGAAGGTACCAAGGCGATCGCATCGATCGCGGCGGATATCAATCCC 5001
QY 361 TyrLeuAlaPheSerAlaIleAlaAlaGlyIleSerGlyIleGluGluLysLeuGlu 380
Db 5002 TATCTGGCTTTGCGAGGCTCATCGCGCGGCTGAAGCGCTGGATGAAGCTGAA 5061
QY 381 LeuProProLaserGlyAsnValTyrAsnAspLysGluLeuProGluPheProAsn 400
Db 5062 CTGACGAACTTTTGGCGCGATGCTATAGCGCGGTGAGGTGAAGGAATTCCTTAC 5121
QY 401 SerLeuGlnAsnAlaThrHisLeuLeuLysGluSerLysMetLeuAsnLysThrPheGly 420
Db 5122 ACCTGGCGGAGCGGCTGCGGCGCTCAAGGTTCCGCTTCTGGAAGAGGCTTCGGC 5181
QY 421 GluLysLeuLeuLeuHisTyrValAsnAlaAlaAsnValGluLeuAsnGluPheSerLys 440
Db 5182 GAGGATGCTGTAACATTACACCCATACCGCCATTGGGAGCAGATCGAATATGACCGC 5241
QY 441 GlnValThrAspTrpGluLeuAsnGlnGlyPheAsnArgTyr 454
Db 5242 CGCGTGACGACTGGGAGCTCTACCGCGGATTTGAACGCTAC 5283

RESULT 3
SMES91790/c
LOCUS SMES91790 323450 bp DNA linear BCT 05-JUL-2002
DEFINITION Sinorhizobium melliloti 1021 complete chromosome; segment 9/12.
ACCESSION AL591790 AL591688
VERSION AL591790.1 GI:15075230
KEYWORDS
SOURCE Sinorhizobium melliloti (Rhizobium melliloti)
ORGANISM Sinorhizobium melliloti
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
1 (bases 1 to 323450)
Capela,D., Barloy-Hubler,F., Gouzy,J., Bothe,G., Ampe,F., Batut,J.,
Boistard,P., Becker,A., Boutry,M., Cadieu,E., Dreano,S., Gloux,S.,
Godrie,T., Goffeau,A., Kahn,D., Kiss,E., Lelaure,V., Masuy,D.,
Pohl,T., Portecelle,D., Puehler,A., Fumelle,B., Ramsberger,U.,
Renard,C., Thebaud,P., Vandenbol,M., Weidner,S. and Galibert,F.
Analysis of the chromosome sequence of the legume symbiont
Sinorhizobium melliloti strain 1021
Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9877-9882 (2001)
21396507
MEDLINE 2 (bases 1 to 323450)
11481430
PUBLISHED
Gouzy,J.
REFERENCE
Direct Submission
Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELILO
EU Consortium
COMMENT
MELILO EU Consortium:
Laboratoire de Biologie Molculaire des Relations
Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet,
France, Laboratoire de Genetique et Developpement UMR6061-CNRS,
Faculte de Medecine, 2 avenue du Pr. Leon Bernard, F-35043 Rennes,
France, GATC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany,
Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25,
D-33615 Bielefeld, Germany, Unite de Biochimie physiologique,
Universite Catholique de Louvain, Place Croix du Sud 2, Bte 20,
B-1348 Louvain-la-Neuve, Belgium, Unite de Microbiologie, Faculte
des Sciences Agronomiques de Gembloux, Avenue Marechal Juin 6,
B-5030 Gembloux, Belgium. E-mail:Jerome.Gouzy@coulouise.inra.fr
http://sequence.toulouse.inra.fr/melliloti.html.
Location/Qualifiers
1 . 323450
/organism="Sinorhizobium melliloti"
/mol_type="genomic DNA"
/strain="1021"
/db_xref="taxon:382"
155. .1519
/gene="SMC01575"

FEATURES
source
155. .1519
/gene="SMC01575"
/function="miscellaneous; hypothetical/partial homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/translate=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC46788.1"
/db_xref="GI:15075231"
/db_xref="SPTREMBL:Q92K34"
/translation="MTPATRTSLSLPAALAMELAPANAQDAYGGYGGGDMVLVT
PGSDLDYIPGAEVHAMDRRRTVDFPMGNIVATVVPNDGQGRQRREYGRDGY
PEPRDRGYSGEPFTGAIPYRDIAPAPEREDLPNSLSDREEAAYDPQDDP
LAQPPAMPVTGKSAEIALQVFLDRGFSFVIDGKMGNSVTKAIEAWQATGET
LQNPNTDLERLRFNGGLPIITTYITADAAGPFVASIPDYAHKAQLPHLSFTVST
EMLGEKFMDEAVLRELNGVDFSPGTTIKVNPNGNKKGVARIVADKARKQVLAY
DRAGKLIAYPSTIGSSDTPSPSGTVHVERIAFDPGTYTNPKINFGQANDRLIQLOP
GNGPVGTVTIVIALSKPTYGTHGTPSPSKIGTKQSHGCVRLTNWDTELGLMWSVGTVV
EFVD"
1707. .2996
/gene="SMC01576"
/707. .2996
/gene="SMC01576"
/function="small molecule metabolism"
/note="Product confidence : putative
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/translate=11
/product="PUTATIVE OXIDOREDUCTASE PROTEIN"
/protein_id="CAC46788.1"
/db_xref="GI:15075232"
/db_xref="SPTREMBL:Q92NJ3"
/translation="WATERALPNLWHATAPARTAPLAGDLTVEIVAIVGGGTGLSA
ALHAEFTGIRTAIEARMIGFCGSRNVLGNAGMWQDDLIATIAAAGNRLLDEL
GGGPFVYDLVAKHGIECAVRNGTLHMSVGAELKEIRERAKMKRGAPVEVLSAE
KAHLSGAREGFTGALLDRAGTIQPLAYARGLARALAAGAEITDTPILLAASRQDGL
HNKTPRGTVTARHVILATNAYGSLVTGVPWKEYRQELTILPYFPQATNLPDVAAR
ILPERQGAWDTGLVMTSFMRQNRNLIFGSIQRDLAIAAGTHRAFAARSVRKLPFYIG
DFRFBHWDGRIGMTNNLPAHVLPANVVISGYNVGIAPGTVEGRALARHVTGDT
SAIPLAETPTVPDPWRTLKSAFYHAGAQAQKHFIDKRF"
complement (3024. .3740)
/gene="SMC01579"
complement (3024. .3740)
/gene="SMC01579"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/translate=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAC46789.1"
/db_xref="GI:15075233"
/db_xref="SPTREMBL:Q92NJ2"
/translation="MAVITFANAKGAGKTTAALILSTELARQGNRVVVLADAPQRWI
TWSVSGRVANLEVISHTVPASLPCHIRELGEADFIVIDLAGAKDAIVALAIGLSD
HVLIPVQSCANDARGAVQILELRHIGEKARVINHSVLTWNISLVTTALQTIKAL
LASRGVSLDTPVERVAREIFECEGGTILQWMDPNRVSNLDKARENAYALAAEVQNL
PVTARRALMSRLRSALPRAA"
complement (3949. .5094)
/gene="SMC01582"
/gene
```

```
CDS
complement(3949..5094)
/gene="SMc01582"
/EC number="1.1.1.1"
/function="small molecule metabolism"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=1
/product="PUTATIVE ALCOHOL DEHYDROGENASE PROTEIN"
/protein_id="CAC46790.1"
/db_xref="GI:15075234"
/db_xref="SPTREMBL:Q92N19"
/translation="MTITANWSYPTAVKFGAGRIKELADHCKALGKKPLVLTDRGLA
PMAITQALDILEAGGIGRAIPADVDPNDPNDLEAGVKAFRDGDHGVAFGGSGSL
DLGCVAFMAGCTPVDVEDIGIDWTRASVEGIAPIVAVPTTAGTSEGVASPVITN
SASHVKVIFHPKPLPGVTICDPELTVMGPVKVITAGTGMDFAFHCLSAYSPPFYHPMS
AGIALGEMRLVKYLPKAYKDGDALEARNMSAAAGAVAFQKGLGAIHLSHPVGA
IYNTHGWTNAVMPVLRFNRSIAEIKIGRAAYLGIAGGDFGDFYDVLRLREELGV
PKLSALGVGTDRIDEMAELVDPDTAGGNPVELTDLAAEKLFAECI"
complement(5091..6476)
/gene="SMc01588"
/EC number="1.2.1.-"
/function="small molecule metabolism"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=1
/product="PUTATIVE ALDEHYDE DEHYDROGENASE PROTEIN"
/protein_id="CAC46791.1"
/db_xref="GI:15075235"
/db_xref="SPTREMBL:Q92N10"
/translation="MTMIRCVSPVDGVEYAEERPAIPLMARQAVAHARLAQKAWARP
LDERVLVLGAVAKLNEHVEVPELAWQRPVRYGGFEGFNERSNYVASTAADAL
KPLVESDFERIAEPHGVFVIAWPNYPYMTAINTVAPALMAGNTVILKHASOT
ILVGERMVRAFAGVPADVFQNLFDHDTTAALIAAKSPDFINFTGSGVGRSIERA
AAGFTGLGLGGKDPGYMEDADLDAVDTLMDGATYNSGCCGGERIYVHESLY
DAFEKSVAVSNVYKLNPLDPEITLGPMAKPEATVRNQVADAIKGRALIDPKL
FPQDDGAYLAPQVLYVDHSMFMEETFPAPVGMKVSADAEALIMDSRSLGLTA
SLWTKDAERAIRGRDLETGTVMFNADYLDPALCWTGVKETGRGSLSVLGFHNLTR
PKSYHLKKKVT"
complement(6674..8038)
/gene="SMc01594"
/complement(6674..8038)
/gene="SMc01594"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=1
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAC46792.1"
/db_xref="GI:15075236"
/db_xref="SPTREMBL:Q92N19"
/translation="MSYSFEELKEDVAAGRIDTVLACQVDMQRLMGKRPHAFYFVES
AWKETHSCNTLLADMETVPGVKAATSWKGYDYTMKPDLSLTLRIPWLEGTALVL
CDMLDHDTHAEVPHSPAILKQVAREANGKAYMASELEFFLPQSYDARLSGVR
DLQLASYNEDYHI FQTTKEEDVNRALINGLQAGIPVENSKEASQGEINRYRAD
ALTMADHRAIKNGCKEIAWORGKAITFLAKWNYSAAGSSSHIHSLWSMDGETPLFF
DKNGQYQMSLMRHYVAGQLAHASEVTYFLAPYINSYKRFMAGTFAPTKAIWSDKDRNT
```

```
AGYRLCGESKAIIRIECRVCGSDLNPLYLFAALIAAGIENKMELEAPFVGDAYQG
KEVREIPHTLRREAGEALSGKMLRAARGEVVDHYHAAEWEEQOYDVRVTDWEVARG
FERA"
complement(8131..9678)
/gene="SMc01597"
/complement(8131..9678)
/gene="SMc01597"
/function="cell processes; transport of small molecules;
Alignment Scores:
Pred. No.: 2,25e-77 Length: 323450
Score: 963.00 Matches: 198
Percent Similarity: 62.91% Conservatives: 87
Best Local Similarity: 43.71% Mismatches: 168
Query Match: 40.53% Indels: 0
DB: 1 Gaps: 0
US-10-098-602a-2 (1-454) x SME591790 (1-323450)
Qy 1 MetThrIleThrTyrAspGluLeuAsnAsnLeuIleAargAsnGlyLysIleAspThrVal 20
Db 8038 ATGAGCTATTTCGTTTCGAGGAACGTGAAAGAGAGATGTTCCCGCGCGCGCATCGACACGGTG 7979
Qy 21 ValLeuAlaCysValAspMetGlnGlyAargLeuMetGlyLysAargLeuThrGlyArgHis 40
Db 7978 CTCGCCCTGTTCAGGTGGACATGCGAGCGCGCTGATGGCGCAAGCGTTCATCGCGCAATAC 7919
Qy 41 PheLeuGlyLeuAspGlnLysLysIleSerIleSerThrPheValTyrAlaValThrIle 60
Db 7918 TTCTGGAAAGCGGTGGAAGGAAACCATAGCTGCAACTACCTGCTTGGACCGGACATG 7859
Qy 61 GluGlyIleAlaGlyGlyTyrGluIleSerSerValAspThrGlyTyrSerAspCys 80
Db 7858 GAGATCGAGACCGTCCCGCGCTACAGCGCGAGCTGGGGAAGGTTACGGCGACTAC 7799
Qy 81 HisLeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTyrSerGluGlyAlaValLeu 100
Db 7798 ACGATGAAGCCGACCTTTCGAGCGCTCCGCGCATTCGTTGGTTCGAAGGACGCGCGCTG 7739
Qy 101 AlaIleSerAsnProHisAsnPheValThrSerGluProLeuPheCysSerProArgVal 120
Db 7738 GTTCTGCGCATCATGCTGGACCATGACACGCGATGCGGAGTGGCGCATTCGCGCGCGCC 7679
Qy 121 IleLeuMetGlnGlnIleGluAargLeuAlaAsnLeuLysLysGlyLeuPheAlaSer 140
Db 7678 ATCTGAAGAGCAGGTGGCGCGCTCGAAGCCATCGGCTTCAAGGCTATATGCGGAGC 7619
Qy 141 GluLeuGluPheAsnLeuPheAsnGluThrTyrTyrSerAlaSerGlnLysHisTrpLys 160
Db 7618 GAGCTCGAATTCCTCTCTTCGATCATGCTCATGCGATGCGGCGCTCTCAGGATATCGC 7559
Qy 161 AsnLeuLysThrAlaGlnProHisHisGlnTrpMetAsnIleSerAlaSerSerGlyIle 180
Db 7558 GACCTGCGCTGCGCAGCGGTTCACAGAGGACTACACATCTTCCAGACGCAAGAA 7499
Qy 181 GluThrPheMetArgSerValAargAsnLysLeuGluAlaGlyIleLeuMetGluAla 200
Db 7498 GAGATCTGATGCGGCGGATCCGCAACCGGCTTCAGGAGCGCGGATCCCGGTCGAAAC 7439
Qy 201 ThrHisProGluPheLeuProSerGlnHisGluLeuAsnPheValProAlaAspProLeu 220
Db 7438 TCCAAAGGCGCGGCTCCGCGCGGAGGAGATCAATGCTTCGTTACGCGCGCGCTG 7379
Qy 221 ThrMetAlaAspArgHisIleIleAlaLysHisGlyValAargGluMetAlaGluGlnSer 240
Db 7378 ACATGCGCGCAGCGGATGCGATCATCAAGAACGCGTCAAGAACGCGTTCGCTGCGCAGC 7319
Qy 241 GlyMetValAlaThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHis 260
Db 7318 GCGAAGCGGATCACCTTCCTCGCGCAAGTGAACACTCTCGGCGCGCGGCTCTCTGCTCCAC 7259
Qy 261 IleHisMetSerLeuGlnAspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAsp 280
```

Db 7258 ATCCACCAGTCGCTCTGGAGCAAGGACGGCGAAACCGCGCTGTCTTCCGACAGAACGGG 7199

Qy 281 GluTyrGlyMetSerThrLeuAlaAraGlnTyrPheAlaGlyLeuLeuLeuTyrValPro 300

Db 7198 CAGTACGGCATGTCGAGCTCATGCGCCATATGTGCGCGGTGAGTTCGCGCCATCGGAGC 7139

Qy 301 GluAlaThrTyrPhePheAlaSerTyrIleAsnSerTyrIleArgLeuGlnProLeuThr 320

Db 7138 GAGGTACCATTTCTCTGGCGCCCTATCACTCTACAAAGCGCTTCATGGCGCGTACC 7079

Qy 321 PheAlaProThrLysCysTyrPheAlaIleAsnArgThrSerAlaPheArgLeuCys 340

Db 7078 TTGCTCCGACCAAGCGATCTGGAGCAAGCAACCGCCGCGCTATCGCTCTGCTG 7019

Qy 341 AsnSerLysSerGluGlyIleAsnValGluLeuArgIleGlyAlaPheLeuAsnPro 360

Db 7018 GCGAGGCGGCAAGCAATCCGATCGAATCGCGGTGCGCGCTCCGATCTCAATCCG 6959

Qy 361 TyrLeuAlaPheSerAlaIleIleAlaAlaGlyIleSerGlyIleGluGluLysLeuGlu 380

Db 6958 TATCTTGCTTTCGCGCCCTGATCGCGCGCGGATTCGCGCATCGAACAAGATGGA 6899

Qy 381 LeuProProAlaSerGlyAsnValTyrAsnAspLysGluLeuProGluPheProAsn 400

Db 6898 CTGAGGCGCGTTCGTGGCGGACGCTATCAGGCAAGGAGGTGCGCGAGATCCGCGAT 6839

Qy 401 SerLeuGlnAsnAlaThrHisLeuLeuLysGluSerLysMetLeuAsnLysThrPheGly 420

Db 6838 ACCTGAGAGAGCGCGCGGAGCGCTCTCGCGCTCGAAGATGTCGCGCGCTTTCGCG 6779

Qy 421 GluLysLeuLeuHisTyrValAsnAlaAlaAsnValGluIleAsnGluPheSerLys 440

Db 6778 GAAGAGTCTGATACCATATGTGCAGCGCGAGATGGGAGCAGGATATGACCG 6719

Qy 441 GlnValThrAspTrpGluLeuAsnGlnGlyPheAsnArg 453

Db 6718 CCGGTGACCATTTGGAGGTGCGCGCGGTTCGAAAGG 6680

RESULT 4

AE009691/c 12340 bp DNA linear BCT 20-MAR-2003

LOCUS Brucella melitensis 16M chromosome II, section 53 of 107 of the complete sequence.

ACCESSION AE009691

VERSION AE008918

KEYWORDS AE009691.1 GI:17984724

SOURCE Brucella melitensis 16M

ORGANISM Brucella melitensis 16M

REFERENCE 1 (bases 1 to 12340)

AUTHORS DelVecchio,V.G., Kapatral,V., Redkar,R.J., Patra,G., Mujer,C., Los,T., Ivanova,N., Anderson,I., Bhattacharyya,A., Lykidis,A., Reznik,G., Jablonaki,L., Larsen,N., D'Souza,M., Bernal,A., Mazur,M., Goltzman,E., Selkov,E., Elzer,P.H., Hagius,S., O'Callaghan,D., Letesson,J.-J., Haselkorn,R., Kyrpides,N. and Overbeek,R.

TITLE The genome sequence of the facultative intracellular pathogen Brucella melitensis

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (1), 443-448 (2002)

PUBLISHED 11/5/6688

REFERENCE 2 (bases 1 to 12340)

AUTHORS DelVecchio,V.G., Redkar,R.J., Patra,G. and Mujer,C.

TITLE Direct Submission

JOURNAL Submitted (13-NOV-2001) Institute of Molecular Biology and Medicine, University of Scranton, Scranton, PA 18510, USA

REFERENCE 3 (bases 1 to 12340)

AUTHORS Elzer,P.H. and Hagius,S.

TITLE Direct Submission

JOURNAL Submitted (13-NOV-2001) Department of Veterinary Science, LSU Ag Center, 111 Dairymple Building, Baton Rouge, LA 70803, USA

REFERENCE 4 (bases 1 to 12340)

AUTHORS Kapatral,V., Los,T., Ivanova,N., Anderson,I., Bhattacharyya,A.,

Lykidis,A., Reznik,G., Jablonaki,L., Larsen,N., D'Souza,M., Bernal,A., Mazur,M., Goltzman,E., Selkov,E., Haselkorn,R., Kyrpides,N. and Overbeek,R.

Direct Submission

Submitted (13-NOV-2001) Integrated Genomics, Inc., 2201 W. Campbell Park Drive, IL 60612, USA

REFERENCE 5 (bases 1 to 12340)

AUTHORS Letesson,J.-J.

TITLE Direct Submission

JOURNAL Submitted (13-NOV-2001) Unite de Recherche en Biologie Moleculaire, Laboratoire d'Immunologie et de Microbiologie, Universite of Namur, 61 rue de Bruxelles, Namur 5000, Belgium

REFERENCE 6 (bases 1 to 12340)

AUTHORS O'Callaghan,D.

TITLE Direct Submission

JOURNAL Submitted (13-NOV-2001) Faculte de Medecine, INSERM U431, Avenue Kennedy, Nimes 30900, France

FEATURES

source

1..12340

/organism="Brucella melitensis 16M"

/mol\_type="genomic DNA"

/strain="16M"

/db\_xref="taxon:224914"

/chromosome="II"

/complement(235..678)

/gene="BMEI10552"

/complement(235..678)

/gene="BMEI10552"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAL53794.1"

/db\_xref="GI:17984725"

/translation="MNSFRKTCAGALALIFGATSIPTVAAPMNDPAINONVIOAR AHYRQNTNRGRPGYHGRGIRYHGRHNRDNGWYPLAAGAGAILGAIISQPR PVYRAGAPSHVQWCYSRYKSYRASNDTFQPYNGPKRCRSPYSR"

/complement(860..2029)

/gene="BMEI10553"

/complement(860..2029)

/gene="BMEI10553"

/EC\_number="1.1.1.1"

/codon\_start=1

/transl\_table=11

/product="ALCOHOL DEHYDROGENASE"

/protein\_id="AAL53795.1"

/db\_xref="GI:17984726"

/translation="MTLITAKMNPFTVLFGCHIKELPAVLKAAGIEHPFTDPLG AKLPVYASTLAVLDAAGVQYGVSDVKPNPVSNLHAGVKVFRGHGDIAPFGGSA LDGLKLI AFQAGQTRPVMDFEDIGDMTRADSNAPVIAVPTTAGTSEVGRAGVLT NESTHTKVIHPKMLPVTIADPELSTGMPPTTAGTMDALAHCLAYCAPGFHPM ADGIAVEGIRLVFENLPKAYANGDI EARNMMAAAMGATAFOKGLGAIHSLSPHG ALYDTHGNTNAPVFPVYVHNRAAIEKRIARLADYLGIEGGDFADAIKLRALG VRTUPDLIKLQMONAKALAEAMAVADFTAGGNPVELTKDALKLNALEGR"

/complement(2130..3434)

/gene="BMEI10554"

/complement(2130..3434)

/gene="BMEI10554"

/EC\_number="6.3.1.2"

/codon\_start=1

/transl\_table=11

/product="GLUTAMINE SYNTHETASE"

/protein\_id="AAL53796.1"

/db\_xref="GI:17984727"

/translation="MLACIVDQGLIKGRFYGFQFVSGYDTHGCVLLADDIDME PVFGYAAAGWDKGYGDFVNMKPLSTLRAPLWLEKTAI VLCDVDFHHDDLSHPRAVL KKQVORLHERGYRAYFASLEFYI FDETYKSAKRKHMETASPVQCYVHLITRE EPLRAMRNHLADAGIPVENSGSGQQLNRYCKALENDRHVIMKNAMKEIAE AHGKCTIFMAKYDARAGSSSHVNSIWSADCKPLFPDKAPYTMPLMRSWAQOI KYATDYTVLAPYINYSYKRFQAGTAPTKIMSDNRRTAGFRLCGEGTKGRIEGRIG KADINPYLAPALINAGLKVDEKLELDEPFVFGDAYSAVRLKEIPYTLREAAALGKS AFLKPAFGEDVNNHYTHAHWEQIEYDRKRVTDWELYRGFERI"

/complement(3548..4315)

/gene="BMEI10555"

gene

CDS









```

/codon_start=1
/transl_table=11
/product="AGR C 3880p"
/protein_id="AAK87886.1"
/db_xref="GI:15157278"
/translation="MTVIONISPIDGSAAREAMSLEAARAAVSTARRAQKWARPP
LEDVQLVILGVARLKNEMAEVPELAHMMGRPYRGTEFFGRNERNYVASIAADAL
APLVIESGNFERREAREAGVFEVIAPWNPYNTAINTVAPALMAGNTYVRIKHAQT
LLGGERVRAPEVGVPEVDFINFLDHDGTTSTLISEGLFNFNFTSGSVEGGRAIERA
AAGTFTGLLEGLGKDCPYVMEDADDAVDTLMDGGTYNSGCCGCCGIERIYVNNELY
DEFVDEKSAVVAPOILLVNDHSEMEFTEETFGPAIGIMVKVNDDEALAMANDSKYGLT
ASLTQDAARARIGRIEIGTVFNRADYLDLPALCWTVKVKETGCGGSLSVLGFQNLIT
RPSYHLKKVTK"
complement(3888..4334)
/gene="AGR C 3881"
CDS
complement(3888..4334)
/gene="AGR C 3881"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR C 3881p"
/protein_id="AAK87887.1"
/db_xref="GI:15157279"
/translation="MILNGKSSMTSDGFLDTCVAVWMSGEPVSDVSAALNQSVR
AGDPVCVSAVTAMGMMLVAKRISETKSPQRWYDDFKREAVIEQPTVDFIATSCF
LPQVHKDPTDRILITITAREHDLITITRDRVILAYGAGHVKTILAC"
complement(4297..4725)
/gene="AGR C 3882"
CDS
complement(4297..4725)
/gene="AGR C 3882"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR C 3882p"
/protein_id="AAK87888.1"
/db_xref="GI:15157280"
/translation="MANTGQDIKSWFSQOLKRKGKQADLAFNLNDRSAVTRMLNG
DRMSVEQDRIAEYLFIPVGDVALHRRGVAGVSENNOYATSEPPAPSGRSPEAGNV
STYESRPIFGCKMGTITVMPDVLTKPVDVFEWGEKLYNE"
complement(4788..6152)
/gene="AGR C 3883"
CDS
complement(4788..6152)
/gene="AGR C 3883"
/notes="probable glnA4 protein - Mycobacterium tuberculosis
(strain H37Rv)"
/codon_start=1
/transl_table=11
/product="AGR C 3883p"
/protein_id="AAK87889.1"
/db_xref="GI:15157281"
/translation="MTSYTFDALKMDVAEGRIDTVLACLVDMDQRLMGKRFQAEFFVE
SAPETHSCNYLATDEMETVPGYKSSNEKGYDYLKPDILSTLKRKFWLPGTALV
LCDVLDHHTHEEPHSPRALLKQVAREAMGLKAYNATELEFFLFDQTPDARASGY
KDLNLSAGNEDHYI FOTTKEEDVMRALRGLQAGIPVENSKEGASPGQAEINRYIA
EALTMDADRAIIKNATIEAMSKGKAVTFLAKWNNAAAGSSSHIHSGLWDGKPAFL
DKDGEKMSDMVHRHYVAGLLHASDITYFLAPYINSYKRPWAGTFAPTKAIWSLDNPT
AGVRLCCAEKTKGRIEIRCVGSDNPLVANAALLAGIDGKLELEPFAVSDAVGG
KDVREIPKTLRDATAFDLGSKMLRQAFEDDDVDVHYVHAARWEQEEDRRVTDNEVARG
FERA"
complement(6218..7774)
/gene="AGR C 3885"
CDS
complement(6218..7774)
/gene="AGR C 3885"
/notes="probable transport protein PA4023 {imported} -
Pseudomonas aeruginosa (strain PAO1)"
/codon_start=1
/transl_table=11
/product="AGR C 3885p"
/protein_id="AAK87890.1"
/db_xref="GI:15157282"
/translation="MADWLVFLLSGEKGMDSSSGVSYYKKADASYFEKRLGSLRYAGVW
SLMALGVAVISGHFSGWNFSTGGWGMVLVAGIIIAIMYLGITFSLAEMSPALPHT

```

```

GAAYSPARTAMGPMGQFVTGLCENVEYVLTAVVVTFTAYVNSILGLDPAYSFPVMI
VFYALFALNFGLEISFKVTLVITLISLAVVFFVITSAIPNIDFSFALNIGVGPDC
KAVELPEGGSGFFPGFGLATLPPAVMLFLAIQLPLAABESVDKRWMPKGIILG
MVTLMVSAFMIVLLNPSLPGVGAFLHSSLSLEPLDGGFKAIYDGGVGLLGLVALTGLI
ASPTLILYAGROGRIYSLSRAGYPTVLSITHSKYRTPYVANIITGAIYGLAVMLVWFS
LGAEOGSGIISGVLNNAVEGAMFSYIMQAIISFILLKKNLPNIERPRSPGPIGALL
TIIIAIVTLIYIQDNPFTKGVLMVAVFAVAFAYFAVGVGRHRLILSPSEBEFALEHKQ
AAVAAAAKA"
complement(7956..8729)
/gene="AGR C 3887"
CDS
complement(7956..8729)
/gene="AGR C 3887"
/notes="similar to orf3 gene in Methylobacterium
extorquens"
/codon_start=1
/transl_table=11
/product="AGR C 3887p"
/protein_id="AAK87891.1"
/db_xref="GI:15157283"
/translation="MTVRSRFFTEAEGKAVGVENAAAKGDVLLVCEHASATIPQKYGT
LGLSADVLSHAAWDPGALAVARLLSEKPHATLVYQFSLRVYDCNRPSPSPAMPVK
SEIYDIPGNFDLDEAREFARTSALYVPFHDVSEIIAEROAAGRKVVVVITHSPTPV
HGRPREVEIGILHNDLSRLADAMLAGAEGASLTVRNDPYGPDGVTHTLRLHALPDG
LUNVMEIRNDLIANEQEQAIAAGFLHELMKALSSIEE"
8704..9780
/gene="AGR C 3888"
CDS
8704..9780
/gene="AGR C 3888"
/notes="hypothetical protein PA5506 {imported} -
Pseudomonas aeruginosa (strain PAO1)"
/codon_start=1
/transl_table=11
/product="AGR C 3888p"
/protein_id="AAK87892.1"
/db_xref="GI:15157284"
/translation="MKNRERTVMDPARKPAIPMTKRIILHIIASVKNLKFSLHFSI
DSAAGTDVYFSAPEADRRKSKLSRTATVSDVIHAHYDALTRSEKRLAESLLGNYPV
SNAGSITTTAENAGVSTPVARMVQKLYGYPEFOAHLHOLEATISGSPVAKHNRWA
TNAPGILHNRPADATIGNRLDGLDITAVFDNAARLLDKRKSIVFVCGITGALA
EXFTTHQVIRPKTTLMSNSSSAWPOYMLNMSAGDVLVIFDIRYBHDMTTLAEVAKA
NGVQIILFTDQWTSVPARHALTFVRVRIEAPSAWSDSVVTLFVVEALIEAVQNGTWDE
TKERNALGELFEQTRLFRRPERT"
complement(9799..10044)
/gene="AGR C 3889"
CDS
complement(9799..10044)
/gene="AGR C 3889"
/notes="hypothetical protein"

```

## Alignment Scores:

Pred. No.:	2,01e-77	Length:	11253
Score:	942.50	Matches:	195
Percent Similarity:	63.11%	Conservative:	89
Best Local Similarity:	43.33%	Mismatches:	165
Query Match:	39.67%	Indels:	1
DB:	1	Gaps:	1

US-10-098-602a-2 (1-454) x AE008130 (1-11253)

Qy	4	ThrTyAspGluLeuAenAenLeuLeuAraGAsGlyLysIleAaspThrValValLeuAala 23
Db	6140	ACATTCGACGACCTATAAATGGATGTCGCCGAGGGGGCGATCGACATGCTTCTGCCGTCG 6081
Qy	24	CysValAspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHisPheLeuGly 43
Db	6080	CTTGTGCACATGAGGGGGCTCTGATGGGCAACGGTTCACGGCGGAATTCCTTCGTTGAA 6021
Qy	44	LeuAspGlnLysIleSerIleSerThrPheValAlaValThrIleGluGlyIle 63
Db	6020	AGCGCCTTTGAGGAACCCATAGCTCAACTATGTCTGCTCCACCCGACATGGAGATGAA 5961
Qy	64	AlaGlyGlyGlyTyrGluIleSerValAspThrGlyTyrSerAspCysHisLeuCys 83
Db	5960	ACCGTTCCCGGTTACACGCTCGAGTGGGAAAGGATATGCGGATATATACGCTGAAG 5901

QY 84 AlaAspLeuAsnSerLeuHisLeuLeuProTropSerGluGlyAlaValLeuAlaIleSer 103  
 Db 5900 CCGGACCTTTCCGAGCTGCGGAAGTGGCGTGGCTGCGAGGTACGGCGTGGTTTGTGC 5841  
 QY 104 AsnProHisAsnPheValThrSerGluProLeuPheCysSerProArgValIleLeuMet 123  
 Db 5840 GAGTGCTGCACCATCACACCCATGAAGAAGTCCCGCATTCGCCGCGCGCTCTCGAA 5781  
 QY 124 GlnGlnIleGluArgLeuAlaAsnLeuLysLeuLysGlyLeuPheAlaSerGluLeuGlu 143  
 Db 5780 AAGCAGGTGGCGGCTGCGAGCGCATGCGGCTGAAGCCCTATATGCGCAGCACTAGAA 5721  
 QY 144 PheAsnLeuPheAsnGluThrTyrlsSerAlaSerGlnLysHisTrpLysAsnLeuLys 163  
 Db 5720 TTCTTTCTGTTCCAGCAGACTTTCGATGCGGCGCGCCGAGCGGTTTATAAGACCTCAAT 5661  
 QY 164 ThrAlaGlnProHisHisGlnTrpMetAsnIleSerAlaSerSerGlyIleGluThrPhe 183  
 Db 5660 CTCGCCAGCGGTATATGAGGATTACCATATCTTCCAGCAGCAAGGAGGATGTC 5601  
 QY 184 MetArgSerValArgAsnLysLeuGluGluAlaGlyIleLeuMetGluAlaThrHisPro 203  
 Db 5600 ATCGCGGCGTGGCAAGGCGCTGCAGGCTGCGGAATTCGCTCGAGAATTCGAAGGCG 5541  
 QY 204 GluPheLeuProSerGlnHisGluLeuAsnPheValProAlaAspProLeuThrMetAla 223  
 Db 5540 GAGGCGTCTCCCGTCCAGCGGAAATCAATGTGCGTTATGCCGAGCGCTGACCATGGCT 5481  
 QY 224 AspArgHisIleAlaLysHisGlyValArgGluMetAlaGluGlnSerGlyMetVal 243  
 Db 5480 GACCGCAGCCCATCATCAAGACGCCACCAAGGAGATGCTGTCGAAGGCGCAAGGCA 5421  
 QY 244 AlaThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHisMet 263  
 Db 5420 GTCACTGCTCTGCCAAGTGAATCAATGCGCGCGGAGTTCATCGCATATCCACAG 5361  
 QY 264 SerLeuGlnAspAlaGluThrGluLysAsnAlaPheTyrlsAspGlnAsnAspGluTyrls 283  
 Db 5360 TGCTCTGGAGCGTGGAC--GGTAAGCGCGGCTCTCTCGACAGGATGCGCAACATGGC 5304  
 QY 284 MetSerThrLeuAlaArgAsnTrpIleAlaGlyLeuLysTyrlsValProGluAlaThr 303  
 Db 5303 ATGTCGAGTGTGATGCGGCATATGTTGCGGCTTCTGGCGCATCCGACGACATCACC 5244  
 QY 304 TyrPhePheAlaSerTyrlsAsnSerTyrlsArgLeuGlnProLeuThrPheAlaPro 323  
 Db 5243 TATTTCTGGCGCATACATCAATCTCTACAGCGCTTCATGCTGGCAGCTTTGGCGCG 5184  
 QY 324 ThrLysCysCysTrpAlaIleAspAsnArgThrSerAlaPheArgLeuCysAsnSerLys 343  
 Db 5183 ACAAGCGCATCTGGAGCTCGACACCGCAGCGAGGATACCGCTTTGCGCGCGGAA 5124  
 QY 344 SerGluGlyIleAsnValGluLeuArgIleGlyAlaAspLeuAsnProTyrlsLeuAla 363  
 Db 5123 ACCAAGGCAATCGTATCGATGCTCGCGCGGCTCGACCTCAATCCCTATCTCGCC 5064  
 QY 364 PheSerAlaIleIleAlaGlyIleSerGlyIleGluGlyLysLeuGluLeuProPro 383  
 Db 5063 ATGGCGCTCTCTTCGAGCGGCAATGACGGTATCGAAATAAATCGAATTCGAGCC 5004  
 QY 384 ProAlaSerGlyAsnValTyrlsAsnAspLysGluLeuProGluPheProAsnSerLeuGln 403  
 Db 5003 GCTTTTGTGCTGATGCTATGCGGTAAAGGAGTGGCGGAGATACCAAGACATTCACG 4944  
 QY 404 AsnAlaThrHisLeuLeuLysGluSerLysMetLeuAsnLysThrPheGlyGluLysLeu 423  
 Db 4943 GAGCGCAGCGCTTTCTGGACGATCGAAATGTCTGCGCAGCGCTTTGGCGAGCAGTG 4884  
 QY 424 IleLeuHisTyrlsValAsnAlaAlaAsnValGluIleAsnGluPheSerLysGlnValThr 443  
 Db 4883 GTCGATCACTACGTGATGCGCAAGGAGTGGGAGCAGGAAGATACGACCGCGCGTGACC 4824

QY 444 RepTtpGluLeuAsnGlnGlyPheAsnArg 453  
 Db 4823 GACTGGAGTGGCAGCGGTTCGAGCGA 4794  
 RESULT 7  
 LOCUS MTW003 13246 bp DNA linear BCT 02-SEP-2002  
 DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 125/162.  
 ACCESSION AL008883 AL123456  
 VERSION AL008883.1 GI:3261490  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mycobacterium tuberculosis H37Rv  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
 tuberculosis complex.  
 REFERENCE  
 AUTHORS  
 Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,  
 Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E.,  
 Tekala, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,  
 Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,  
 Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,  
 Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,  
 Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,  
 Squares, S., Squares, R., Sultston, J.E., Taylor, K., Whitehead, S. and  
 Barrell, B.G.  
 Deciphering the biology of Mycobacterium tuberculosis from the  
 complete genome sequence  
 Nature 393 (6685), 537-544 (1998)  
 JOURNAL  
 MEDLINE 98295987  
 PUBMED 9634230  
 REFERENCE  
 AUTHORS  
 Parkhill, J.  
 Direct Submission  
 Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium  
 tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome  
 Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique  
 Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,  
 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk  
 On Jun 27, 1998 this sequence version replaced gi:2612801.  
 NOTES:  
 Details of M. tuberculosis sequencing at the Sanger Centre are  
 available on the World Wide Web.  
 (URL, [http://www.sanger.ac.uk/Projects/M\\_tuberculosis/](http://www.sanger.ac.uk/Projects/M_tuberculosis/)) CDS have  
 been renumbered from the original cosmid submissions but the old  
 gene designations are in brackets after the new gene numbers.  
 Gene prediction was based on a Hidden Markov Model of TB genes  
 implemented in TPAse (Krogh) supplemented with visual inspection  
 of positional base preference in codons, especially where there is  
 an increase in the observed/expected third position G + C.  
 CAUTION: In some cases we may not have predicted the correct  
 initiation codon. Where possible we choose an initiation codon  
 (atg, gtg, or ttg) which is preceded by an upstream ribosome  
 binding site sequence (optimally 5-13bp before the initiation  
 codon) if this cannot be identified we choose the most upstream  
 initiation codon.  
 FEATURES  
 Location/Qualifiers  
 1..13246  
 /organism="Mycobacterium tuberculosis H37Rv"  
 /mol\_type="genomic DNA"  
 /strain="H37Rv"  
 /db\_xref="taxon:83332"  
 <1..>13246  
 /note="fragment designated v003. Does not represent a  
 physical clone"  
 complement(60..836)  
 /gene="RV2857C"  
 complement(60..836)  
 /gene="RV2857C"  
 /note="RV2857C"  
 /note="RV2857C", (MTW003.03c), len: 258 aa. Probable  
 dehydrogenase, similar to many eg. P94681 toluenesulfonate  
 zinc-independent alcohol dehydrogenase (tsac) from  
 Comamonas testosteroni (252 aa), fasta scores; opt: 515

z-score: 632.1 E(): 4.7e-28, 38.0% identity in 250 aa overlap. Also similar to many M. tuberculosis dehydrogenases eg. MTCY39.16c (38.3%identity in 248 aa overlap)"

/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein Rv2857c"  
/protein\_id="CAA15519.1"  
/db\_xref="GI:2612802"  
/gb\_xref="SPTREMBL:O33339"  
/translation="MMDLSORLAGRVAIVITGGSGIGLAAGRRMRBEGATVVGVDVDF  
EAGGNAADELSGLVFPTDVCDDAVNGLVFGAASYTGRIDIAFNAGISPPENLIEN  
TELAQRQVDNLKSVLGQICRAALRHMLAKGSI VNTAS FVAMGSATFSQISYTAS  
KGVLNARELGVOFARQIRVNALCPGVNTPLLQLFAKNPERAARMVMHVPLQRF  
APEPILAAAFVFLASDDASFITASFTLVGDGISSAYVTPL"  
complement(833..2200)  
/gene="aldC"  
complement(833..2200)

CDS  
/gene="aldC"  
/note="RV2858c, (MTW003.04c), len: 455 aa. aldC,  
Probable aldehyde dehydrogenase, aldC; similar to many eg.  
FEAB ECOLI P80668 phenylacetaldhyde dehydrogenase (499  
aa) \_fasta scores; opt: 1074 z-score: 1239.2 E(): 0, 42.2%  
identity in462 aa overlap. Contains PS00687 Aldehyde  
dehydrogenases glutamic acid active site and PS00070  
Aldehyde dehydrogenases cysteine active site. Also similar  
to many M. tuberculosis dehydrogenases eg. MTCY369.13  
(38.1% identity in 467 aa overlap)"  
/codon\_start=1  
/transl\_table=11  
/product="aldC"  
/protein\_id="CAA15520.1"  
/db\_xref="GI:2612803"  
/gb\_xref="SPTREMBL:O33340"  
/translation="MSUTQLNPATESVLASVDHTDANAVDDAVQARAARQWRALA  
PAQRAAGURFAAAVOHLDLEALEAVNSGHPIVSAEWAGHRVDVLATYASAPETLP  
SGROI PVAGVGVDTFNEPMGVGVITPNFPMWIASWAIPALGAAGNAVILPKAEPLTP  
LTQVRGLAVAEAGDELDDLQPGKGTVGGERVTHPDIKI VFTGSTEVGRKVMAG  
AAAOKRLTVBELGKGSANI VFPHDCDLERAATAPAGVFDNAGODCCARSRLIVQSRVY  
DRFWELLPEPVHSIVVGDGPGRATGMGPLVSRAHRDKVAGVPDPDAVPRGTAPAGR  
DFPEPTPTLVTRGDRVTVDIFGPVVVVLTFDDEADAISLANDTAYGLSGSIWTDDLI  
SRALKARAVASGNLSVNSVSRFRNTPFFGGFKOSGVRELGPAPLQFTETKNVFLA  
VGEEM"

RBS  
complement(839..843)  
/gene="aldC"  
/note="possible RBS upstream of Rv2857c"  
complement(1400..1435)

misc\_feature  
/gene="aldC"  
/note="PS00070 Aldehyde dehydrogenases cysteine active  
site"  
complement(1496..1519)  
/gene="aldC"  
/note="PS00687 Aldehyde dehydrogenases glutamic acid  
active site"  
complement(2197..3123)  
/gene="Rv2859c"  
complement(2197..3123)  
/gene="Rv2859c"  
/note="RV2859c, (MTW003.05c), len: 308 aa. Unknown;  
similar to E. coli hypothetical gene TR:P76038  
(EMBL:AB000228) (258 aa), fasta scores; opt: 326 z-score:  
354.5 E(): 1.4e-12, 30.7% identity in 238 aa overlap.  
Contains three 17 aa repeats at the N-terminus very  
similar to those in other M. tuberculosis proteins e.g.  
Y130 MYCU Q10699 putative DNA polymerase cy49.30 (393:  
aa)"

misc\_feature  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein Rv2859c"  
/protein\_id="CAA15521.1"  
/db\_xref="GI:2612804"  
/gb\_xref="SPTREMBL:O33341"  
/translation="MDLSASRSDDGDLPLRPASPRLRSPVSDGDPRLPASPRLRSPVSP

```
Db 4453 TTGCGGTGGACCGAGTTGGAGCGACTGGTCCGCGCGGTGACGTCGACACCGCTCATGCTC 4394
Qy 23 AlaCysValAspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHisPheLeu 42
Db 4393 GCCTTACCAGCATGCGAGCGCGGTGGCGGCAACCGATATCGGCGCGCATTTTCGTC 4334
Qy 43 ---GlyLeuAspGlnLysIleSerIleSerThrPheValTyrAlaValThrIleGlu 61
Db 4333 GACGACATACCCACCGCGCGTGGAGTCTGTCAGTTATCTGTCGCGGTGGACGTCGAC 4274
Qy 62 GlyIleAlaGlyGlyTyrGluIleSerSerValAspThrGlyTyrSerAspCysHis 81
Db 4273 CTGAACACGCTGCCGCTATGCGATGCGCAGTTGGACACCGGCTACGCGCATGGTG 4214
Qy 82 LeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTyrPheSerGluGlyAlaValLeuAla 101
Db 4213 ATGACGCGGACATGTCCTCATCTGCGGCTGATTCCTTGGTACCGGAACGGCGCTGGTG 4154
Qy 102 IleSerAsnProHisAsnPheValThrSerGluProLeuPheCysSerProArgValIle 121
Db 4153 ATCGCGAC---CTGGTCTGGCGCGGACGAGTCTCGCTCTCGCGCGCAGCAT 4097
Qy 122 LeuMetGlnGlnIleGluArgLeuAlaAsnLeuLysLeuLysGlyLeu-----Phe 138
Db 4096 CTGCGCGCTCAGCTCGATCGG-----CTCAAGCGCGCGGACTGGTTCGCCGATGTG 4046
Qy 139 AlaSerGluLeuGluPheAsnLeuPheAsnGluThrTyrLysSerAlaSerGlnLysHis 158
Db 4045 GCACCGAGTGGAGTTCATCGTGTTCGACCGCTATCGCCAGGATCGGCGACGCGCGG 3986
Qy 159 TrpLysAsnLeuLysThrAlaGlnProHisGlnTrpMetAsnIleSerAlaSerSer 178
Db 3985 TATCGCGGTGACCCCGCGGCGGACGACTACACATCGACTACCGATATGGCATCTCTCG 3926
Qy 179 GlyIleGluThrPheMetArgSerValArgAsnLysLeuGluGluAlaGlyIleLeuMet 198
Db 3925 CGGATGAGCGGTTCGTCGCGCATCGGTTGGTATGGCGGTGGTCTCGCATTC 3866
Qy 199 GluAlaThrHisProGluPheLeuProSerGlnHisGlnTrpMetAsnLeuPheValProAlaAsp 218
Db 3865 GAGCGGTCAAGCGGATCAACATCGGCGGACGAGGATCGGTTTCGTTACGACGAG 3806
Qy 219 ProLeuThrMetAlaAspArgHisIleAlaLysHisGlyValArgGluMetAlaGlu 238
Db 3805 CGCGTGTCTACCTCGCACCACTCGCATCTACAGACGCGCCCAAGGAATCGCGAC 3746
Qy 239 GlnSerGlyMetValAlaThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAla 258
Db 3745 CAGCAGCGCAAGAGCCTTACGTTTCATGCGCGAATACGATCAACGCGAA---GGTAATAGC 3689
Qy 259 CysHisIleHisMetSerLeuGlnAspAlaGluThrGluLysAsnAlaPheTyrAspGln 278
Db 3688 TGTCAATCATCTGCTGCTGCGTGGCGGACGATGGTTCGCG---GTGTTTGGCGACAGT 3632
Qy 279 AsnAspGluTyrGlyMetSerThrLeuAlaAsnTrpIleAlaGlyLeuLeuLysTyr 298
Db 3631 AACGGCGCGCACGCGCATGCTCGATGTTCCGCGAGTTCGTCGCGCGCAGTTGGCCACG 3572
Qy 299 ValProGluAlaThrTyrPhePheAlaSerTyrIleAsnSerTyrIysArgLeuGlnPro 318
Db 3571 TTGCGCGGAATTCACGCTGTCTATGCGCGCACCACTTAACCTCTACAAGCGATTTGCCGAT 3512
Qy 319 LeuThrPheAlaProThrLysCysTyrTrpAlaIleAspAsnArgThrSerAlaPheArg 338
Db 3511 AGCAGTTTCGCGCGGCGGCTGGTGGTGGGCTGGACAAATCGACCTCGCGCCCTGCGG 3452
Qy 339 LeuCysAsnSerLysSerGluGlyIleAsnValIleLeuArgIleGlyGlyAlaAspLeu 358
Db 3451 GTGTTT---GGCCACGGCAAAACATCCGGTTCGAATCGCGGTTCGCCGCGGTGATGC 3395
Qy 359 AsnProTyrLeuAlaPheSerAlaIleAlaAlaGlyIleSerGlyIleGluGluLys 378
Db 3394 AACCACTACCTGGCGGTGGCTCTCTATGCTGGAGGGTGTACGATATCGAGCGGGC 3335
```

```
Qy 379 LeuGluLeuProProAlaSerGlyAsnValTyrAsnAspLysGluLeuProGluPhe 398
Db 3334 CTTACGTCGCCGAGCCCTGTCTCGCAACGCTACCAAGCGCCGATCGAAGCGCTG 3275
Qy 399 ProAsnSerLeuGlnAsnAlaThrHisLeuLeuLysGluSerLysMetLeuAsnLysThr 418
Db 3274 CCGGTTACGTCGCCGACGCCGCGGTCTGTTCGAGGATTCGCGCTGGTGGCGAGCG 3215
Qy 419 PheGlyGluLeuLeuIleLeuHisTyrValAsnAlaAlaAsnValGluIleAsnGluPhe 438
Db 3214 TTGCGCGAGGATGTTGTCTCGCACTACCTGGAACACGCGCTGTGGAGCTGGCGCGCTC 3155
Qy 439 SerLysGlnValThrAspTrpGluLeuAsnGlnGlyPheAsnArg 453
Db 3154 AACCGCGCGGTACCGATTGGAGAGGATACGTGGATTGAGCGC 3110

RESULT 8
AE007117/c 14715 bp DNA linear BCT 27-APR-2001
LOCUS Mycobacterium tuberculosis CDC1551, section 203 of 280 of the
DEFINITION complete genome.
ACCESSION AE007117 AE000516
VERSION AE007117.1 GI:13882699
KEYWORDS
SOURCE Mycobacterium tuberculosis CDC1551
ORGANISM Mycobacterium tuberculosis CDC1551
Bacteria; Actinobacteris; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
REFERENCE 1 (bases 1 to 14715)
AUTHORS Fieischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,
Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,
Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,
Gill,J., Mikula,A. and Bishai,W.
TITLE Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 14715)
AUTHORS Fieischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,
Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,
Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,
Gill,J., Mikula,A. and Bishai,W.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1. .14715
/organism="Mycobacterium tuberculosis CDC1551"
/mol_type="genomic DNA"
/db_xref="taxon:83331"
/note="clinical strain"
complement(148..924)
/gene="WT2925"
complement(148..924)
/gene="WT2925"
/note="similar to GB:AE000512; identified by sequence
similarity; putative"
/codon_start=11
/product="oxidoreductase, short-chain
dehydrogenase/reductase family"
/protein_id="AAK47250.1"
/db_xref="GI:13882700"
/transeq="NM150R1AGRVAVITGGSGIGLAAGRMRAEGATIVVGDV
EAGAAADSLGLFVPTDVCDEAVNGLFDGAAETVGRIDIAFNAGISPPENLLEN
TELAARQVQDYNLKSIVLCCRAALRHVYLAGKSVNTFVAVMGSAQSISYAS
KGVLAWSRELGVQFARQIRVNALCPGVNTPLLQELFAKPERAARMHVPLGRF
ASPDIEAFAVFLASDASFITATFLVDGGISSAYVPL"
complement(921..2288)
gene
```

CDS	/gene="MT2926" complement (921. .2288) /gene="MT2926" /note="similar to PID:2282233; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="aldehyde dehydrogenase family protein" /protein_id="AAK47251.1" /db_xref="GI:13882701" /translation="MSTQLINPATEEVLASVDHADANAVDQVRAAQRWARLA PAQRAGLRPAAPAAHLELALEAVNSGHPIVSAWEAGHVRDVLAFVAAFERLP SGROI PVAGGVDTFTNEPMGVGVITPMNFMVWSAIAAPALAGNAGVLVYKPAELTP LTTMRLGLAVAGLEDLLQVLPCKTVGGERFVTHPDIRKI VFTGSTEYVGRKVMAG AAQVKT VLELGGKSANI VEHDCDLERAAATPAGVDFNAGQDCARSLTLVQSVY RFEWELPEPAVHSIVGDPGSRATEMGPVLSRAHRDKVAGVVPDPAVAPFGTAPAGR GFWFPPTVLT PKRGDRTVDSIFGPVVVVLTFDDDEADAI SLANDTAYGLSGSIWTDLL SRARVAVAGSGLNSVNSGSRFNTFPFGFGKSGVGRGLGDPAPLQFTETKNVFIA VGEEM"	CDS	/gene="MT2930" complement (5570. .6154) /gene="MT2930" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAK47255.1" /db_xref="GI:13882705" /translation="MTETGDMVALRVSDADRNGTMRERLHNAVALGINIDFEQORS RVSPACTRSLDGLVGLDLPREGAIVTSAADRVELRWAGLSLKRHGEIVITRLLALVR LGSITELDLVKARFAPVVVLELDMFGLSLEVRPUNGASISDDVEVYVGSASDRRKDA PAGCTPHVITLGRMVCVGVVVKGPRALLRRHG" 6312. .6506 /gene="MT2931" 6312. .6506 /gene="MT2931" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAK47256.1" /db_xref="GI:13882706" /translation="MSLSNWLROAGRLQLEAQRQRLTAQELREFFASRPDETGAEP DWAHLQVMAESRRGLPAP" 6503. .6883 /gene="MT2932" 6503. .6883 /gene="MT2932" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAK47257.1" /db_xref="GI:13882707" /translation="MIFVDVTNFMVAVGRDPLRMPAREFLEHSLHQDLRLVTSAEAM QELNAYVPVGRNSTLDSALTVALTEIWPVEAADVAHARTLHRRHFGLGARDLLHL ACQQRGVTRIKTFDHTLASAFRS" complement (6965. .8821) /gene="MT2933" complement (6965. .8821) /gene="MT2933" /note="similar to GP:1666696; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="penicillin-binding protein, putative" /protein_id="AAK47258.1" /db_xref="GI:13882708" /translation="MSRPVSPQFLTSATVMTKTLASATSGLLLLAVAMSGCTPRP QGPAAEKFPFALAIGDTSAAQLSDNPNEAREALNAWAGLQAHLDAQVLSAKYA EDTGTVAVRSWHLPKDRITWYDYGQLKMARDEGRHWVWRTTSGLHPKLGHQTFALRA DPRRASVSVGGTDVLPGYLYHSLDAGQAGRELFGTAHVVAGLHPFDDTLNDPO LLAEASSTSPDLDTLTHADDSNRVAAAIQOLPGVITTPQALLPTDKRHPFAPVLD VKAIVDELDKAGRWVSVYNGQVDVSVLHEVAPSPASSVITLDRVQNAQAQAVN TRGKAMIVVVKPSTGEILAIQONAGADAPVATGLYPGSGTFKMITAGAAVERDL ATPFTLLGCGEIDIGHRTIPNYGGEDLVVMSRAPASSCNTTFAELSRLLPRGLT QAARVIGILDYQDGIITTVGSPVPTDLAERTEDGFGQVLASFPFGNALVATVA AGKTPVQLIAGRPATVEGDATPI SQMIDALRPMMLRVVTNGTAKETAGCGEVFGKT GEAEFPFGGSHWVAGYRGDLAFALIVGGSGSEYAVRMTKVMFESLPPGYLA" complement (9006. .9716) /gene="MT2934" complement (9006. .9716) /gene="MT2934" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAK47259.1" /db_xref="GI:13882709" /translation="MPPHEHRNPEVLFRSFRSRPADRSQSWLPDSGRAPVGBELLSAV DVRADVDAQDHYRCVLVYVQVAVSRAERAGQLAPKGLAHPOQLARQIAEREPD HCRDSRWQLVEVSTRGCGPHGVRSVSGASRDABFGADLVFAVGAAGNVGVGFS
gene	/gene="MT2927" complement (2285. .3085) /gene="MT2927" /note="similar to GP:3581870; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="glutamine amidotransferase, putative" /protein_id="AAK47252.1" /db_xref="GI:13882702" /translation="MSGDGLPLRPASPLRSLPGASRPVVGTLTAYLEQVTRVWDIPA GYLPADYPEGITMAGGAVALLPPOVPDPESVGCVLDSLHALVITGGVDLPAAVGOEP HPATDHPRGDAWEFALLRGALGMPEVLGICRGTOVLNVALGCTLHOHLPLDILGHS GHRAGNVFTRLPVHTASGTRELAEIGESADVPCYHQAI DQVGEGLVSAVDVGV EALPLPGDTFVLAVQWHPKSLDLRLFLKALVDAASGYAGRQSAEP" complement (3138. .4511) /gene="MT2928" complement (3138. .4511) /gene="MT2928" /note="similar to SP:P36205 GB:X60160 PID:48181. GB:AE000512; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="glutamine synthetase, putative" /protein_id="AAK47253.1" /db_xref="GI:13882703" /translation="MTGPGSPPLAWTELERLVAAAGVDVTIVATFDMQGRLAGKRISG RHFVDDIATRGVCCSYLLAVVDLNTVPGYAMASWDTGDMWMT PDLSTRLIPWL FGTALVADLVWADGSEVAVSPRSILRRQLDKARGLADVATELEFIVFDQPYRQA WASGYRGLTPASDYNIDYAILASSRMEPLLRDIRLGWAGLRFEAVKGCNMQQEI GFYDEALVTCNDHAIYKNGAKETADQKSLTFMAYKDEREGNSCHIHVSLRGTDI AVFADNSGPHGSMFSRFSVAGQATLREFTLCYAPTINSYKRPADESFPATLAKGL DNRTCALRVGCHGNIRVECRPGVDVNOYLAAALTAGLGIERGEGLUPEPCVGNV YQGADVRLPTVLDAADVLFEDSALVREAFGEDVVAHYLNNARVELAAFNAAVTDWER IRGPERL" complement (4671. .5528) /gene="MT2929" complement (4671. .5528) /gene="MT2929" /note="similar to GB:D26562 SP:P07906 GB:M15106 PID:146727 PID:473823; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="methionine aminopeptidase" /protein_id="AAK47254.1" /db_xref="GI:13882704" /translation="MPSRTALSPGLVSPTRFPVNWIA RPEYVVGKPAAQEGSEPVQTP EVIKMRVAGRIAGALAEAGKAVPGVTTDELDRIAHEYLVDNGAVPSTLGYKGFPK SCCTSLNEIVCHIGIPDSTVITDGI VINDVTAYIGVGHGDTNATFPAGDVADHERLV DRTREATMRAINTVKGRLSVIGRVI ESYANRGYNVRDFTGHGIGTTFHNGLVVL HYQPAVETINQGMQFTTIEPMINLGALDYIEWDGWTVTVTVKORKKTAQFEHTLLVTD TGVRLTCL" complement (5570. .6154)	gene	/gene="MT2930" complement (5570. .6154) /gene="MT2930" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAK47255.1" /db_xref="GI:13882705" /translation="MTETGDMVALRVSDADRNGTMRERLHNAVALGINIDFEQORS RVSPACTRSLDGLVGLDLPREGAIVTSAADRVELRWAGLSLKRHGEIVITRLLALVR LGSITELDLVKARFAPVVVLELDMFGLSLEVRPUNGASISDDVEVYVGSASDRRKDA PAGCTPHVITLGRMVCVGVVVKGPRALLRRHG" 6312. .6506 /gene="MT2931" 6312. .6506 /gene="MT2931" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAK47256.1" /db_xref="GI:13882706" /translation="MSLSNWLROAGRLQLEAQRQRLTAQELREFFASRPDETGAEP DWAHLQVMAESRRGLPAP" 6503. .6883 /gene="MT2932" 6503. .6883 /gene="MT2932" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAK47257.1" /db_xref="GI:13882707" /translation="MIFVDVTNFMVAVGRDPLRMPAREFLEHSLHQDLRLVTSAEAM QELNAYVPVGRNSTLDSALTVALTEIWPVEAADVAHARTLHRRHFGLGARDLLHL ACQQRGVTRIKTFDHTLASAFRS" complement (6965. .8821) /gene="MT2933" complement (6965. .8821) /gene="MT2933" /note="similar to GP:1666696; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="penicillin-binding protein, putative" /protein_id="AAK47258.1" /db_xref="GI:13882708" /translation="MSRPVSPQFLTSATVMTKTLASATSGLLLLAVAMSGCTPRP QGPAAEKFPFALAIGDTSAAQLSDNPNEAREALNAWAGLQAHLDAQVLSAKYA EDTGTVAVRSWHLPKDRITWYDYGQLKMARDEGRHWVWRTTSGLHPKLGHQTFALRA DPRRASVSVGGTDVLPGYLYHSLDAGQAGRELFGTAHVVAGLHPFDDTLNDPO LLAEASSTSPDLDTLTHADDSNRVAAAIQOLPGVITTPQALLPTDKRHPFAPVLD VKAIVDELDKAGRWVSVYNGQVDVSVLHEVAPSPASSVITLDRVQNAQAQAVN TRGKAMIVVVKPSTGEILAIQONAGADAPVATGLYPGSGTFKMITAGAAVERDL ATPFTLLGCGEIDIGHRTIPNYGGEDLVVMSRAPASSCNTTFAELSRLLPRGLT QAARVIGILDYQDGIITTVGSPVPTDLAERTEDGFGQVLASFPFGNALVATVA AGKTPVQLIAGRPATVEGDATPI SQMIDALRPMMLRVVTNGTAKETAGCGEVFGKT GEAEFPFGGSHWVAGYRGDLAFALIVGGSGSEYAVRMTKVMFESLPPGYLA" complement (9006. .9716) /gene="MT2934" complement (9006. .9716) /gene="MT2934" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAK47259.1" /db_xref="GI:13882709" /translation="MPPHEHRNPEVLFRSFRSRPADRSQSWLPDSGRAPVGBELLSAV DVRADVDAQDHYRCVLVYVQVAVSRAERAGQLAPKGLAHPOQLARQIAEREPD HCRDSRWQLVEVSTRGCGPHGVRSVSGASRDABFGADLVFAVGAAGNVGVGFS



CDS

complement(99..875)  
/locus\_tag="Mb2882c"  
/EC\_number="1.1.1.1"  
/note="Mb2882c", len: 258 aa. Equivalent to Rv2857c.  
len: 258 aa, from Mycobacterium tuberculosis strain H37Rv,  
(100.0% identity in 258 aa overlap). Probable short-chain  
dehydrogenase/reductase (EC 1.1.1.1), highly similar to  
various dehydrogenases e.g. O88068|SCI35.33c PROBABLE  
DEHYDROGENASE (SDR FAMILY) from Streptomyces coelicolor  
(260 aa), FASTA scores: opt: 1208, E(): 2e-68, (72.35%  
identity in 253 aa overlap); O91376|PA1649 from  
Pseudomonas aeruginosa PROBABLE SHORT-CHAIN DEHYDROGENASE  
(253 aa), FASTA scores: opt: 569, E(): 2.1e-28, (39.2%  
identity in 255 aa overlap); O9EX74|MLHA SDR-LIKE ENZYME  
from Rhodococcus erythropolis (246 aa), FASTA scores: opt:  
567, E(): 2.8e-28, (41.15% identity in 248 aa overlap);  
etc. Also similar to many Mycobacterium tuberculosis  
dehydrogenases e.g. PABG3|Rv2002|MT2058|MTCY39.16c  
PUTATIVE OXIDOREDUCTASE (260 aa), FASTA score: (38.3%  
identity in 248 aa overlap). BELONGS TO THE SHORT-CHAIN  
DEHYDROGENASES/REDUCTASES (SDR) FAMILY."  
/codon\_start=1  
/transl\_table=11  
/product="PROBABLE SHORT-CHAIN TYPE  
DEHYDROGENASE/REDUCTASE"  
/protein\_id="CAD96569.1"  
/db\_xref="GI:31619629"  
/translation="MMDLSQRLAGRVAVITGGSGIGLAAGRMRARGATIVGVDV  
ENGAADELISGLFVPTDVCDEDAVNGLFGCAASTYGRIDIAFNAGISPEDNLIN  
TELAWQVQDNLUKSYLCCRAALRHVLGKSIWNTASFAVMGASISQISYAS  
KGVLANSLRELGVQFARIGRVINCPGVPNTPLQLFAPKNPARRMWHVPLGRF  
APEDEIAAVALFASDASFTTASTFLVDGGSISAVYTPL"  
/gene="aldC"  
/locus\_tag="Mb2883c"  
complement(872..2239)  
/gene="aldC"  
/locus\_tag="Mb2883c"  
/EC\_number="1.2.1.3"  
/note="Mb2883c, aldC, len: 455 aa. Equivalent to Rv2858c,  
(100.0% identity in 455 aa overlap). Probable aldC.  
aldC dehydrogenase (EC 1.2.1.3), similar to many e.g.  
O88069|SCI35.34c PUTATIVE ALDEHYDE DEHYDROGENASE from  
Streptomyces coelicolor (483 aa), FASTA scores: opt: 1872,  
E(): 6.4e-109, (64.5% identity in 448 aa overlap);  
O9FAB1|ALDH|BT-ALDH ALDEHYDE DEHYDROGENASE from Bacillus  
thermoovorans (497 aa), FASTA scores: opt: 1157, E():  
2.1e-64, (44.3% identity in 458 aa overlap); O33455|CYMC  
P-CUMIC ALDEHYDE DEHYDROGENASE from Pseudomonas putida  
(494 aa), FASTA scores: opt: 1149, E(): 6.5e-64, (43.15%  
identity in 452 aa overlap);  
P40047|DHAS YEAST|ALD5|ALDH5|ALD3|YER073W ALDEHYDE  
DEHYDROGENASE from Saccharomyces cerevisiae (Baker's  
yeast) (519 aa), FASTA scores: opt: 1091, E(): 2.7e-60,  
(38.55% identity in 459 aa overlap);  
P80668|FEAB ECOLI|PADA|MAOB|B1385 PHENYLACETALDEHYDE  
DEHYDROGENASE (EC 1.2.1.39) from Escherichia coli strain  
K12 (499 aa), FASTA scores: opt: 1074, E(): 3e-59, (42.2%  
identity in 462 aa overlap); etc. Also similar to many M.  
tuberculosis dehydrogenases e.g. P71823|Rv0768|MTCY369.13  
(489 aa), FASTA score: (38.1% identity in 467 aa overlap).  
Contains P500687 Aldehyde dehydrogenases glutamic acid  
active site and P500070 Aldehyde dehydrogenases cysteine  
active site. BELONGS TO THE ALDEHYDE DEHYDROGENASES  
FAMILY."  
/codon\_start=1  
/transl\_table=11  
/product="PROBABLE ALDEHYDE DEHYDROGENASE ALDC"  
/protein\_id="CAD96570.1"  
/db\_xref="GI:31619630"  
/translation="MSTQLINPATTEEVLSVDHTDANAVDDAVQRAAARWARLRA  
PAQRAAGLRAFAAAVQAHDLAALAEVANSCHPTVSAEWAGHVRDVLAFYAAASPERL

gene

CDS

SGRIIPVAGGVDTFNEPMGVGVITPMFPMVIASWAIAPALAAAGNAVILVKPAELTP  
LTTMRLGLAVEAGDEDLQVLPGKTVVGRFVTHPIRKIVITGSTRVGRKVMAG  
AAQVKRVTLELGKGSANI VFHDCDLERAATTAAGVDFNAGODCCASRTLVQRVSVY  
DRMELELPAVHSIVGDCGRATEMGPVLSRAHRDKAVGVDPDPAVPRGTA PAGR  
GFWPPFVLTPKGRDVTVDIEIFGPVVVLTTFDEADAISLANDATAYGLSGSIWTDLL  
SRALURAVAVESGNLSVSSSRVFNTPFGGFKQSGVGRBELGDPAPLOFTETKNVFIA  
VGSEM"  
complement(2236..3162)  
/locus\_tag="Mb2884c"  
complement(2236..3162)  
/locus\_tag="Mb2884c"  
/EC\_number="2.3.1.5"  
/note="Mb2884c", len: 308 aa. Equivalent to Rv2859c,  
(100.0% identity in 308 aa overlap). Possible  
amidotransferase (EC 6.3.5.- or 2.-.-.-), equivalent (but  
longer 58 aa) to Q9CBU9|MLI573 POSSIBLE AMIDOTRANSFERASE  
from Mycobacterium leprae (249 aa), FASTA scores: opt:  
1226, E(): 3e-64, (71.55% identity in 239 aa overlap).  
Also similar to other amidotransferases and hypothetical  
proteins, but shorter in N-terminus e.g. O88072|SCI35.37  
HYPOTHETICAL 25.3 KDA PROTEIN from Streptomyces coelicolor  
(242 aa), FASTA scores: opt: 683, E(): 1.2e-32, (47.65%  
identity in 235 aa overlap); AAK79730|Q97188|CAC1764  
PREDICTED GLUTAMINE AMIDOTRANSFERASE from Clostridium  
acetobutylicum (241 aa), FASTA scores: opt: 458, E():  
1.6e-19, (32.95% identity in 246 aa overlap);  
AAK75091|Q97QV9|SP1089 GLUTAMINE AMIDOTRANSFERASE CLASS I  
from Streptococcus pneumoniae (229 aa), FASTA scores: opt:  
431, E(): 5.6e-18, (34.75% identity in 236 aa overlap);  
etc. Contains three 17 aa repeats at the N-terminus very  
similar to those in other Mycobacterium tuberculosis  
proteins e.g. Q10699|YY30 MYCTU|Rv2090|MT2151|MTCY49.30  
PUTATIVE 5'-3' EXONUCLEASE RV2090 (EC 3.1.11.-)."  
/codon\_start=1  
/transl\_table=11  
/product="POSSIBLE AMIDOTRANSFERASE"  
/protein\_id="CAD96571.1"  
/db\_xref="GI:31619631"  
/translation="MDLSASRSGDGDPRLPASPRLSRPSVSDGGDPLRAPSPRLSRPSVS  
DGGDPLRAPSPRLSRPSVGLTAYLEQVTVGWDIVPAGYLPADYFEGITPMAGS  
VALLPPQVPDPESVGVLDLSHALVITGGYDLDPAAYGQEPHPATDHPGPRDAPWF  
ALLRQALQMPGVLCRGTVLNVALGTLHQLPDLILGSHGRAGNVFTRLPVHT  
ASTCELAELIGESADVPCHVHQALDQVGEGLVSAVDVGVIEALELPDGTFFLAVQM  
HPEKSLDLRLFLKALVDASGYAGRSQAEPR"  
complement(3143..4516)  
/gene="glnA4"  
/locus\_tag="Mb2885c"  
complement(3143..4516)  
/gene="glnA4"  
/locus\_tag="Mb2885c"  
/EC\_number="6.3.1.2"  
/note="Mb2885c, glnA4, len: 457 aa. Equivalent to Rv2860c,  
len: 457 aa, from Mycobacterium tuberculosis strain H37Rv,  
(100.0% identity in 457 aa overlap). Probable glnA4,  
glutamine synthetase class II (EC 6.3.1.2), similar to  
many glutamine synthetases e.g. O88070|SCI35.35c from  
Streptomyces coelicolor (462 aa), FASTA scores: opt: 1947,  
E(): 8.2e-120, (64.15% identity in 452 aa overlap);  
Q98H15|MLL3074 from Rhizobium loti (Mesorhizobium loti)  
(465 aa), FASTA scores: opt: 1321, E(): 7.8e-79, (46.7%  
identity in 452 aa overlap); O98EM0|MLL4187 from Rhizobium  
loti (Mesorhizobium loti) (456 aa), FASTA scores: opt:  
699, E(): 4.6e-38, (33.5% identity in 454 aa overlap);  
Q9CDL9|GLNA from Lactococcus lactis (subsp. lactis)  
(Streptococcus lactis) (446 aa), FASTA scores: opt: 633,  
E(): 8.2e-34, (32.45% identity in 456 aa overlap); etc.  
Also similar to three other potential glutamine synthetases  
in Mycobacterium tuberculosis:  
Q10378|GLN2 MYCTU|GLNA2|Rv2222c|MT2280|MTCY190.33c|MTCY427  
.03c PROBABLE GLUTAMINE SYNTHETASE (446 aa), FASTA score:  
(31.1% identity in 453 aa overlap); Rv1878|glnA3 and



## JOURNAL

Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces sequencing team, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk  
On or before Oct 29, 2002 this sequence version replaced gi:20520886, gi:20520900, gi:20520927, gi:20520938, gi:20520939, gi:20520834, gi:20520914, gi:20520916.

## COMMENT

## FEATURES

## source

Location/Qualifiers

1..277000  
/organism="Streptomyces coelicolor A3(2)"  
/mol\_type="genomic DNA"

/strain="A3(2)"

/db\_xref="taxon:100226"

226..993

/gene="SC01445"

/note="synonym: SCL6.02"

226..993

/gene="SC01445"

/note="SCL6.02, hypothetical protein, len: aa; similar to various hypothetical proteins, e.g. TR:CBAS2958 (EMBL:AL109950) Streptomyces coelicolor hypothetical 25.2 kD protein, SCJ4.24c, 242 aa; fasta scores: opt: 424 z-score: 482.7 E(): 1.8e-19; 39.9% identity in 218 aa overlap and to TR:O9BWF2 (EMBL:AL451182) Streptomyces coelicolor hypothetical protein SCK13.32c, 263 aa; fasta scores: opt: 711 Z-score: 779.3 E(): 8.9e-36; 49.407% identity in 253 aa overlap"

/codon\_start=1

/transl\_table=11

/product="conserved hypothetical protein SCL6.02"

/protein\_id="CAB76867.1"

/db\_xref="GI:7209205"

/translation="MTTILVTGGTGLRLVTERLRTGGHEVRVLSRHSFPYAVDLRE GSGDLAALGADVTTHVCAITQGGDSAAALIAARACAGATHLVISIVGVDRVPL GYRSKVAEKLVAESGIGTIVLRATQFDHVLQVOMLAKLPPVVLPAHSDOPVEV AEVADRLEAQAQAGACRVPDMGQPEVTFASLARAYLGATGRRRAVNVVPLAGRTYR AFRAGHLEAHAVEGTFPEYLAGRGHREGRGG"

1064..1432

/gene="SC01446"

/note="synonym: SCL6.03"

1064..1432

/gene="SC01446"

/note="SCL6.03, possible integral membrane protein, len: 122 aa. Contains possible hydrophobic membrane spanning regions"

/codon\_start=1

/transl\_table=11

/product="possible integral membrane protein"

/protein\_id="CAB76868.1"

/db\_xref="GI:7209206"

/translation="MDPAAARAVLEVALCPAALHEHLRLDFGAGLALAVMLCAAV ATWERRLVLPVACPDSVVARLLFHARHLEPLSAGGAAGFMAPLGTAVLLPTVLGWP AADGTAPEPGSAGTGRPG"

complement(1454..2653)

/gene="SC01447"

/note="synonym: SCL6.04c"

complement(1454..2653)

/gene="SC01447"

/note="SCL6.04c, possible ROK-family transcriptional regulatory protein, len: 399 aa; similar to TR:CBAT71832 (EMBL:AL13862) Streptomyces coelicolor putative ROK-family regulatory protein SCB4A.27, 409 aa; fasta scores: opt: 938 z-score: 965.1 E(): 0; 42.5% identity in 398 aa overlap and to SW:XYLR ANATH (EMBL:269782) Anaerocellum thermophilum xylGse repressor XylR, 399 aa; fasta scores: opt: 327 z-score: 341.8 E(): 1.3e-1; 22.2% identity in 397 aa overlap. Contains Pfam match to entry PF00480 ROK, ROK family and possible helix-turn-helix motif at residues 26..47 (+2.84 SD)"

/codon\_start=1

/transl\_table=11

/product="putative ROK-family transcriptional regulatory protein"

/protein\_id="CAB76869.1"

misc\_feature  
/db\_xref="GI:7209207"  
/translation="MPASPSTABAINDLRLALHLIQREGPLTAGOLKLTGLSRPTVAD LVRSLTASGLIEVAGSGEORGNKLYGIVAEAYALADLVTEGVAVLVSDLLGR VLAESVPIIDHSGTGPVQAVSLVERAKEAGADRLHTVGIGAPGLIDPAGGELRD STGLPEWHRLMALQLEKFPPEARVGVENETNLAALAEQDGAARDTFTVLLWLGIGI GAAVLDGLRRGVSGAGIEGMPVGTAGLPSATDCDGGFHSVAGAAVYVLDPGC GVTAPAGPEPFAAALVRRSASGSDPAERFLDALADRVVLGAAVYVLDPGC LVLAGETGRAGADALADRVQHLRTRNSPLATEVRPSTLGGGAVLRGALLTARDRAQDD LFAPPER"

complement(1934..2104)

/gene="SC01447"

/note="Pfam match to entry PF00480 ROK, ROK family, score 45.80, E-value 1.5e-12"

complement(2662..2667)

2753..2757

2766..3977

/gene="SC01448"

/note="synonym: SCL6.05"

2766..3977

/gene="SC01448"

/note="SCL6.05, probable transport protein, len: 403 aa; similar to TR:AAF12043 (EMBL:AE002079) Deinococcus radiodurans transport protein, putative DR9502, 471 aa; fasta scores: opt: 687 z-score: 761.8 E(): 0; 39.0% identity in 359 aa overlap and to SW:MOSC\_RHIME (EMBL:U17071) Rhizobium meliloti membrane protein MosC, 407 aa; fasta scores: opt: 496 z-score: 552.3 E(): 2.4e-23; 31.0% identity in 378 aa overlap. Contains match to Prosite entry PS00217 Sugar transport proteins spanning regions"

/codon\_start=1

/transl\_table=11

/product="putative transport protein"

/protein\_id="CAB76870.1"

/db\_xref="GI:7209208"

/translation="MNTGTGVRREHARYAAVFAVHGSVTGSFATRPVWIOQHAG VSQQLGALATFALGASDVAANLGVETRLDKSINSLGHGMSTGALIGSAAGTLAA TLCLALFVYGATAGMSDVANLGVETRLDKSINSLGHGMSTGALIGSAAGTLAA HLCTDARTHLLAAAVLTVTGLVACTWLDLQPAEDEDPPRFPALPKSALLIGAVGF CAVFAGASLDNSAVYLRDELETSAGLAAACTTGFTLTMARLAGDKVVDPRGAVRT VRVSGVLTGGLLVVGGHPAVMTGFALMGIGIAVVPVPLCAAGAGSNPSLAIA GVATITTSGLVAPSAIGGLAQATSLVWSFGLVTLACCLVFAVRLRAGDRNRPKIS PPVAEVPGRQS"

3093..3170

/gene="SC01448"

/note="PS00217 Sugar transport proteins signature 2"

complement(3937..4395)

/gene="SC01449"

/note="synonym: SCL6.06c"

complement(3937..4395)

/gene="SC01449"

/note="SCL6.06c, hypothetical protein, len: 152 aa; similar to SW:YBGC\_ECOLI (EMBL:M16489) Escherichia coli hypothetical 15.6 kD protein in CydB-TolQ intergenic region ybgC, 134 aa; fasta scores: opt: 140 z-score: 193.8 E(): 0.0023; 26.7% identity in 116 aa overlap"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="CAB76871.1"

/db\_xref="GI:7209209"

/translation="MTAEAPTALSHGRVLPVTVHFDLALGLLHNRAPVYVVERA WTFLWAGHGVREGEDEAGDAGNARELTIGVEAPVTRFCAYAVHLNRLTGTGLT YGRFCSADGATTYARGTRVLRDLARTLRPAWSEAFRTAGRALURPAD"

complement(4404..4407)

4564..6000

/gene="SC01450"

/note="synonym: SCL6.07"

4564..6000

/gene="SC01450"

/note="SCL6.07, probable uracyl permease, len: 478 aa; similar to TR:BAA35783 (EMBL:D90738) Escherichia coli

uracil transport protein Pyrp, 442 aa; fasta scores: opt: 908 z-score: 949.9 E(0): 0; 43.1% identity in 429 aa overlap and to SW:PYRP\_BACCL (EMBL:X76083) Bacillus caldolyticus uracil permease (uracil transporter) Pyrp, 432 aa; fasta scores: opt: 770 z-score: 806.6 E(0): 0; 33.0% identity in 437 aa overlap. Contains Pfam match to entry PF00860 xan ur permease, Xanthine/uracil permeases family. Contains also possible hydrophobic membrane spanning regions"

```
/codon_start=1
/transl_table=11
/product="putative uracil permease"
/protein_id="CAB76872.1"
/db_xref="GI:7209210"
/translation="MDLGVYRWKLHGDKVPAQAVRDPRLSWPETAAGLQHVAVM
FGASFAVPLVGLDPLNLAIMMSGVATVIFLLATRGVPSYLGCSLSFVGVAIVRAQG
GTSATVTGAVLVVGVAVFLGAVRFGARIHAAMPVVTGAVVLMILGFLNLAIPVTS
TYWQDQWTLFVMLTFLGAVCLRGFNSRIIFLGLVFGVLSWALDRIFGKIHSVD
GSGKPTDHWLDFSAVGQADWIGLPSFGPSPENSAIIVLPAVIAENAGHVKA
GEMTGNDLKLGTATASDQIGSVLSTAVGPPENTYSENIGVMAATRVYSTAAVAA
AGPALLFGLCPKFGAVVAALPGVLGGITVILYGMIGLGAQIWINAKVDMRNLNV
PAAAGIIGVGNVSMETDTFSLSGIALGLTVVITGYHALRAPAPAPAHKLTQPLDEG
TTSYDAKTGGSGAGSGDGGGQAQARS"
/misc_feature
4654..5793
/genes="SCO1450"
/note="Pfam match to entry PF00860 xan_ur permease,
```

misc\_feature

#### Alignment Scores:

Pred. No.: 4,78e-62 Length: 277000  
 Score: 798.50 Matches: 176  
 Percent Similarity: 57.52% Conservative: 84  
 Best Local Similarity: 38.94% Mismatches: 177  
 Query Match: 33.61% Indels: 15  
 DB: 1 Gaps: 8

US-10-098-602A-2 (1-454) x SCO939109 (1-277000)

```
QY 6 AspGluLeuAsnAenLysLeuGluValLeuAlaCysVal 25
DB 184054 GAGAACTGACGCCCTCGTGGCGCGGTGACATCGACACCTCGCTCGCTTCC 183995
QY 26 AspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHisPheLeuGlyLeuAsp 45
DB 183994 GACATGAGGGCGCGCTCCAGGGCAAGCGGTTCGCGCGCGGTTCCTCTCT 183944
QY 46 GlnLysLysLysSerLysThr-----PheValTyrAlaValThrLeuGlu 61
DB 183943 GATGAGGTCTTGGAAACACGCGAGGGCTGCAACTACTCTCGCGGTGGACGCGGAC 183884
QY 62 GlyLeuAlaGlyGlyTyrGluLeuSerSerValAspThrGlyTyrSerAspCysHis 81
DB 183883 ATGAACACCGTCACGCGTACGCCATGCTCTCTGGGACCGCGCTACGCGGACTTCGCG 183824
QY 82 LeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTrpSerGluGlyAlaValLeuAla 101
DB 183823 ATCGCGCGCGACCGCGCCACCTCGCGCGCGTCCCTCGAAACGAGGCGACCGCGATGGCC 183764
QY 102 IleSerAsnProHisAsnPheValThrSerGluProLeuPheCysSerProArgValle 121
DB 183763 GTCCGCGGAC---CTCGCTCGGAGGACGCGTCCCGGTGCTCGCGCGCGCGCGCGATC 183707
QY 122 LeuMetGlnGlnLeuGluArgLeuAlaAsnLeuLysLeuLysGlyLeuPheAlaSerGlu 141
DB 183706 CTCCGCGCGCGCGTGGAGCGGCTCGCGCGACCGGTACCCACAGGTTCGCGCGCGGAG 183647
QY 142 LeuGluPheAsnLeuPheAsnGluThrTyrLysSerAlaSerGlnLysHisTrpLysAsn 161
DB 183646 CTGGAGTTTCATGCTCTTCGCGGACACCTACGAACAGCGCTGGGACGCGAACTACCGCGG 183587
QY 162 LeuLysThrAlaGlnProHisHisGlnTrpMetAsnLysSerAlaSerSerGlyLeuGlu 181
DB 183586 CTGACCCCGCGGCAACCACTGACGCGTACTACTCGGTCTCGCGCACCGCGCGCGTGA 183527
```

```
QY 182 ThrPheMetArgSerValArgAsnLysLeuGluAlaGlyLeuMetGluAlaThr 201
DB 183526 CCCCTGCTGGCGCATCGCAACGAGATGGCGCGCGCGCTCACCGCTCGAGTCGCGC 183467
QY 202 HisProGluPheLeuProSerGlnHisGluLeuAsnPheValProAlaAspProLeuThr 221
DB 183466 AAGGCGGAGTGTGAACCCCGCGCAGCAGATCGCTTCCGTACGACAGGCGCTTCGTC 183407
QY 222 MetAlaAspArgHisLeuAlaLysHisGlyValArgGluMetAlaGluGlnSerGly 241
DB 183406 ACTTCGACCAACGACCGCTCTACAGACCGCGCGCGCAAGAGATCCGCCCGCAGGAGG 183347
QY 242 MetValAlaThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisLeu 261
DB 183346 ATGTCCCTCACCTTCATGCCCAAGTACACGAGCTGGAA--GGCAACTCCTGCCACATC 183290
QY 262 HisMetSerLeuGlnAspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAspGlu 281
DB 183289 CACCTCTCGTCCGCGACCGCGAC--GGCGCAACGCCATGCCGAGGAGGG----- 183239
QY 282 TyrGlyMetSerThrLeuAlaArgAsnTrpIleAlaGlyLeuLysTyrValProGlu 301
DB 183238 ---GGAATGTCGACGCTCATCGGCACCTTCCTCGCGGAGACAGCTGTGCGCGCTGCCGAG 183182
QY 302 AlaThrTyrPhePheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPhe 321
DB 183181 TTCTCCCTCTCTACGCGCGCGCACATCAACTCTCTACAGCGCTTCGAGCCGCTTCTTC 183122
QY 322 AlaProThrLysCysCysTrpAlaLeuAspAsnArgThrSerAlaPheArgLeuCysAsn 341
DB 183121 GCGCGCGCGCGCTCGCTGGCGCGCACGACAAACCGCACCTGCGCGCTCGCTGCTGCT 183065
QY 342 SerLysSerGluGlyLeuAsnValGluLeuArgGlyGlyValAspLeuAsnProTyr 361
DB 183064 GCGCGCGCGCTCTCTCGTTTCGAGAACCGCGTCCCGCGGCGGACGCTCAACCCCTAC 183005
QY 362 LeuAlaPheSerAlaIleAlaGlyIleSerGlyIleGluGluLysLeuGluLeu 381
DB 183004 CTCGCGGTGGCGCGCTGTGTCGCGCGCGATTCGACGCGCATCGAGCGCGCTGGAAC 182945
QY 382 ProProAlaSerGlyAsnValTyrAsnAspLysGluLeuProGluPheProAsnSer 401
DB 182944 CCGGAGCGCTGCGCGCGCAACGCTACAC--GCGCACTTCGCGCAGCTCCGCCACACC 182888
QY 402 LeuGlnAsnAlaThrHisLeuLysGluSerLysMetLeuAsnLysThrPheGlyGlu 421
DB 182887 CTGCGGAGCGCGCGAGCTGTGGGAGAACACGACACCTCGCAAGCGCGCTTCGCGGAC 182828
QY 422 LysLeuLeuLeuHisTyrValAsnAlaAlaAsnValGluIleAsnGluPheSerLysGln 441
DB 182827 GAGTCTGTCGCCCATCTACCGGAACATGGCCCGCGTCACTGAGCGCTTCGACGCGCT 182768
QY 442 ValThrAspTrpGluLeuAsnGlnGlyPheAsnArg 453
DB 182767 GTCAACCGACTGGGAACCTGCGCGCTCTTCGGAACGC 182732
RESULT 11
AP005047 298550 bp DNA linear BCT 10-MAY-2003
LOCUS Streptomyces avermitilis genomic DNA, complete genome, section 27/30.
DEFINITION AP005047 BA000030
VERSION AP005047.1 GI:29610176
KEYWORDS Streptomyces avermitilis MA-4680
SOURCE Streptomyces avermitilis MA-4680
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1
AUTHORS Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Oonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M.
TITLE Genome sequence of an industrial microorganism Streptomyces
```

avermittilis: deducing the ability of producing secondary metabolites  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 AUTHORS

Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M. and Omura, S.  
 Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermittilis*  
 Nat. Biotechnol. 21 (5), 526-531 (2003)  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 AUTHORS

3 (bases 1 to 298550)  
 Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kishida, N., Director-General of Biotechnology Center, Shiba, T., Sakaki, Y. and Hattori, M.  
 Direct Submission  
 Submitted (29-MAR-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center; 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan  
 (E-mail: bio@nite.go.jp, URL: http://www.bio.nite.go.jp/, Tel: 81-3-3481-1933, Fax: 81-3-3481-8424)  
 This work was done in collaboration with Haruo Ikeda (\*1), Jun Ishikawa (\*2), Akiharu Hanamoto (\*3), Chigusa Takahashi (\*3), Mayumi Shinose (\*3), Hiroshi Horikawa (\*4), Hidekazu Nakazawa (\*4), Tomomi Osonoe (\*4), Norihiro Kishida (\*4), Hiseashi Kikuchi (\*4), Tadayoshi Shiba (\*5), Yoshiyuki Sakaki (\*6, \*7), Maehira Hattori (\*1, \*7) and Satoshi Omura (\*1, \*3).  
 Final finishing process and all annotation were done by H. Ikeda and J. Ishikawa.  
 \*1 Kitasato Institute for Life Sciences, Kitasato University  
 \*2 National Institute of Infectious Diseases  
 \*3 The Kitasato Institute  
 \*4 National Institute of Technology and Evaluation  
 \*5 School of Science, Kitasato University  
 \*6 Institute of Medical Science, University of Tokyo  
 \*7 RIKEN, Genomic Sciences Center  
 Following url is also available.  
 http://avermittilis.kitasato-u.ac.jp.  
 Location/Qualifiers  
 1. .298550  
 /organism="Streptomyces avermittilis MA-4680"  
 /mol\_type="genomic DNA"  
 /strain="MA-4680"  
 /db\_xref="taxon:227882"  
 /note="This strain is also named as strain: ATCC 31267, NCIMB 12804 or NRRL 8165.-synonym: Streptomyces avermectinius"  
 66. .827  
 /note="SAV6513"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAC74224.1"  
 /db\_xref="GI:29610177"  
 /translation="MCASENTPRAGGKRAPEGVKHAPEPAVTVDLAVLTIRAGVLQ VLVVERQEPYAGRWALPGGFLVPHESAARAEAEETGLADLTGLHLEQLRTYSE PDRPRNRSVATALLPDPPEPHGGDAAQWLRYNALGPIAFDHRILDAHER VDKLEYLTATSCFPPEFTLGELOQVETVWGTALDRPNFRKRLATPGFVEQIPGA ARLTGGKRGKPAALYRAGGATTLHPPLRLPTSEGRP"  
 827. .1849  
 /gene="draG2"  
 827. .1849  
 /gene="draG2"  
 /note="SAV6514"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative ADP-ribosylglycylhydrolase"  
 /protein\_id="BAC74225.1"  
 /db\_xref="GI:29610178"  
 /translation="MTTTLTKRAATGSLILGALGDALGFPTFENDVPSILAKCGPWR

CDS  
 1853. .2602  
 /note="SAV6515"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAC74226.1"  
 /db\_xref="GI:29610179"  
 /translation="MDALDIDILAPVVAQDPVLFATVSGAHLGYGPPSRDSDIDLRG AHLLPAELVGLREPTETSRMDRDGVEMDLVTHDLKRFVRLMLRNGVYVLEQLLSP LVHTSDAHRELAELAPGVLTTRHAAHYRGFATTOMRLFKTGLKPLLYTFRVLLTG IHLRSGEVOAHLPTLVGQVDAPGMLPELLAAKAREHGAADVQDQARVRLDVERLHGL LDEAGASRLPDAPTAYDALHDLVVRRLSG"  
 complement (2528. .3289)  
 /note="SAV6516"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAC74227.1"  
 /db\_xref="GI:29610180"  
 /translation="MPETDPLGRLVRRHHTIYSCVMSRAFGLATDGSOTDRRGVFL APTPLWSPAKPPSHVGEPAEQFSWELERFCBLALRANPNILECLHSLPLVEQADATG REILLAREAFLSQTHETFAVRSORKLEADVRAHGA PRWKHAMHLRLLTSCRDLLRGLSITDVGDRREPLLAVERGEVMPVESRMTLAAETDEAARTPLPAEPDRRR VDFLARTIRHASALQAHDPDEVGVVGRGVGQP"  
 complement (3311. .3775)  
 /note="SAV6517"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative iron sulphur protein"  
 /protein\_id="BAC74228.1"  
 /db\_xref="GI:29610181"  
 /translation="MTIRPTRRLTLLTGGAVALLAGSKYGDSSNSSQSTSGTPSGS PSGTASATAGGAAGPELAKTSDIPVGGGTIFKDEKVVVTVQEQGFKAFAVCTHQ GCTVTVSDGTINCPCHGSKFRITDASVAAGPAPRPLPAEKITVEGNSIRLT"  
 complement (3772. .4356)  
 /note="SAV6518"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAC74229.1"  
 /db\_xref="GI:29610182"  
 /translation="MPVDNNAATQSPFPQAHRGPGAARFLVPALVAVAVVGLGAY GKVHDPAGTAFNLGFSSTGAVKSWLATTAFFALVOLVSALMYGKLPGPSWAPVLH RWSGRVAFVAVPVAVVAVVYALGYQTYEPRVLWHSLILGCFFGAFSAKMLLLRAERLP GWLLPIVGLVFAVLGTIWLTSALMFFRTFGVTT"  
 complement (4395. .5090)  
 /note="SAV6519"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAC74230.1"  
 /db\_xref="GI:29610183"  
 /translation="MAGNSPRGPVWKRSGPLTLLAGLLAVFMLSINAMTGTKSTS IAYCEEPAPATVSPSPRSTPPSPSPSPADYAGRTDDNSAVAVSLRDGKA IYFCDRNKESWLKGDVDDGTMKLTGKQGAELNGKLEKGRIRGTVDLGGTRYGFT ADKALKPSGLYRATASVRGAKIDGGMIVLPGGKQVGLNDRDKGKPOAAPRIDPETGAVT VDGQQLTARPVTP"  
 5301. .5663  
 /gene="aroH"  
 5301. .5663  
 /gene="aroH"  
 /note="SAV6520"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative chorismate mutase"  
 /protein\_id="BAC74231.1"  
 /db\_xref="GI:29610184"

gene  
 CDS

```

/translation="MAVRAVRGAVOLERDAADHMDQVSELTAILERNGLTADLLIS
INFTATPDILHSDFPAAARAKGLIVDVPLICAQELDIKGMAPRVRVLAHIESDKPRAE
IAHYLGAAGALKRDKIAQ"
5660..6754
/genes="tyrA"
5660..6754
/genes="tyrA"
/notes="SAV6521"
/codon_start=1
/transl_table=11
/product="putative prephenate dehydrogenase"
/protein_id="BAC74232.1"
/db_xref="GI:29610185"
/translation="MFTALVIGTGLIGTSAALALAGRGVVVHLADHPQARAAALG
ACTDEAPEKPDVLTVAAPPAHVAATLADAMRGVGLGVLDVAVSKGPRRLEALGL
DRKSMSTYIGSPHMSGREKSGPLAATGDLFEGRPWVLTFTDITDEVNLALVLSHC
RAVVMVMDADADHRAVALVSHMPLVSSMVAARLNAEATAVRLCGQGRDVRVLAAS
DRPMVITILSANPGVPVADLLTVDVSDLDLTVRALRSLSQSSDEAKRAGATGIEDVLR
GNAGQVRVPEKPGSAPRTYETVAVLIDDDPGQLARIFADAGLAGVNIEDVRIEHATGQ
OAGLVOLVPEKPEAPVLSAALRERGWALRQ"
5680..7575
/genes="cmk"
5680..7575
/genes="cmk"
/notes="SAV6522"
/codon_start=1
/transl_table=11
/product="putative cytidylate kinase"
/protein_id="BAC74233.1"
/db_xref="GI:29610186"
/translation="MENGAAARTVPAVIVADIDSGTSGKSTSKAVALKGLSYLDTGA
QYRAITMMVNSGIDITDPTAATAVAGKEIISGTPDSAPTTIVDGTVDVAAPRTQEV
TSKVSASVAPVETRIETELQSLATSANGILVSGRDIGTTLVLPDADLKIELTASPE
APKARSGLKADVNSTREALKRDADSSRKTSPSLAKADDAVEVDTSDLTQQVIE
CVVTLVVEKRAAK"
7533..8249
/notes="SAV6523"
/codon_start=1
/transl_table=11

```

## Alignment Scores:

```

Pred. No.: 1.17e-59 Length: 298550
Score: 773.50 Matches: 175
Percent Similarity: 56.64% Conservative: 81
Best Local Similarity: 38.72% Mismatches: 183
Query Match: 13 Gaps: 8
DB: 1

US-10-098-602A-2 (1-454) x AP005047 (1-298550)

QY 6 AspGluLeuAsnLeuLeuArgAsnGlyLeuIleAspThrValValLeuAlaCysVal 25
Db 238691 GAGGAGCTGACGCCCTCTGCGCGGCGCGGAGATCGACATGTCGCTCGCGTTCCCC 238750

QY 26 AspMetGlnGlyArgLeuMetGlyLeuArgLeuThrGlyArgHisPheLeuGlyLeuAsp 45
Db 238751 GATATCGAAGGGCGCTCCAGGCAAGCGTTCCGCCGCGCTTCTCTC----- 238801

QY 46 GlnLeuGlyIleSerIleSerThr-----PheValTyrAlaValThrIleGlu 61
Db 238802 GACGAGGTGTGAGCAGCGCACGCGGAGGATGCAACTACTGTCGCGCTCGACACCGGAC 238861

QY 62 GlyIleAlaGlyGlyTyrGluIleSerSerValAspThrGlyTyrSerAspCysHis 81
Db 238862 ATGAACACCGTGCACCGCTATGACATGTCTCTCGGACCGCGGTGTACCGGCACTTCGCC 238921

QY 82 LeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTyrSerSerGluGlyAlaValLeuAla 101
Db 238922 ATGCATCCGACCTGAGACGCTGCGCCAGGTGCGGTGAGAACCGCGGCGCGCTGCTG 238981

QY 102 IleSerAsnProHisAsnPheValThrSerGluProLeuPheCysSerProArgValIle 121
Db 238982 ATCGCCGAC---CTCGCTTGGAAACGACGCGATCGCGGTGTGTCGCCGCCCGCCAGATC 239038

```

```

QY 122 LeuMetGlnGlnIleGluArgLeuAlaAsnLeuLeuLeuGlyLeuPheAlaSerGlu 141
Db 239039 CTGCGTCCGAGCTGACCGGCTCGCGCCCTCGGCTACACCGCCAGGTGGCACCGAG 239098

QY 142 LeuGluPheAsnLeuPheAsnGluThrTyrIysSerAlaSerGlnLeuHisTrpIysAsn 161
Db 239099 CTGGAGTTTCATGCTTCAAGGACACCTACGAGCAGCTGGGACCGCGGATACAAGGG 239158

QY 162 LeuLeuThrAlaGlnProHisGlnTrpMetAsnIleSerAlaSerSerGlyIleGlu 181
Db 239159 CTCACACCGGCAACACGATACAACATCGACTACTCGGTGCTCGGACGGGACGGATCGAA 239218

QY 182 ThrPheMetArgSerValArgAsnLysLeuGluAlaGlyIleLeuMetGluAlaThr 201
Db 239219 CCCGTGCTGGCGGATCGGACAGATGAGCGCGGGGCTCACCGTCGAGTGGGCC 239278

QY 202 HisProGluPheLeuProSerGlnHisGluLeuAsnPheValProAlaAspProLeuThr 221
Db 239279 AAGGCGGAGTGCACCCCGGCGAGCAGATCGTGTTCGGGTACGACGAGGCCCTGGTC 239338

QY 222 MetAlaAspArgHisIleAlaLysHisGlyValArgGluMetAlaGluGlnSerGly 241
Db 239339 ACCTCGCAGCAGCAGCGCTCTACAAGACCGCGCCAGGAGATCGCGCCCGAGGAGGC 239398

QY 242 MetValAlaThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIle 261
Db 239399 GGTGTCGATCACCCTTCATGCCCAAGTACACGAGCGCGGAG---GGCAATCTCTGTCACATC 239455

QY 262 HisMetSerLeuGlnAspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAspGlu 281
Db 239456 CACCTGTGCTGGCGGACGCGGAC---GGCACCAACGCGCATGCGCGGCGGACGAGCCGCG 239512

QY 282 TyrGlyMetSerThrLeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyrValProGlu 301
Db 239513 ---GGCATGTGCGACGTGATCGCCACTTCTCGCGCGGCGAGCTCGCGCGTGGCGGAC 239569

QY 302 AlaThrTyrPhePheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPhe 321
Db 239570 TTCTCATTCTGTAGCGGCCCAACATCACTGTCACAGCGGTTCCAGCGCGGCTCTCTTC 239629

QY 322 AlaProThrLysCysCysTrpAlaIleAspAsnArgThrSerAlaPheArgLeuCysAsn 341
Db 239630 GCGCGGACCGCGCTGGCGGTACGACACCGCATGCGCGCTGCGGTGGTCT--- 239686

QY 342 SerLysSerGluGlyIleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsnProTyr 361
Db 239687 GCGCACGCGCGCTCGATCGCTTCGAGAACCGGCTCCCGCGCGGTGACGCTCAACCGCGCAC 239746

QY 362 LeuAlaPheSerAlaIleAlaAlaGlyIleSerGlyIleGluGluLysLeuGluLeu 381
Db 239747 CTCGCGGTGCGCGACTGTCGCGCGCGGCTGTACGCGATCGACACAGCTGAGGTG 239806

QY 382 ProProProAlaSerGlyAsnValTyrAsnAspLysGluLeuProGluPheProAsnSer 401
Db 239807 CCGGAGGCTGCGCGGCAACGCTTAC---GCCGCGAGTAGACGACGCTCCCCACCACC 239863

QY 402 LeuGlnAsnAlaThrHisIleLeuLysGluSerLysMetLeuAsnLysThrPheGlyGlu 421
Db 239864 CTGCGGAGGCGCGCAACTCTGGGAGAACACGCCCATCGCCATCGCAAGCGCGCTTCGCGGAC 239923

QY 422 LysLeuIleLeuHisTyrValAsnAlaAlaAsnValGluIleAsnGluPheSerLysGln 441
Db 239924 GACGTGTGCGCCACTACGCAACATGCGCGCGCTCGAAGTGAAGGCTTCGACCGCGG 239983

QY 442 ValThrAspTrpGluLeuAsnGlnGlyPheAsnArg 453
Db 239984 GTGACCGACTGGAGCTGCGCGCTCTTTCGAACGC 240019

RESULT 12
AX259711
LOCUS
DEFINITION Sequence 6 from Patent WO0173038.

```

AX259711 1860 bp DNA linear PAT 26-OCT-2001  
 DEFINITION Sequence 6 from Patent WO0173038.







REFERENCE 2 (bases 1 to 128136)  
 AUTHORS Larbig,K.D., Christmann,A., Johann,A., Hartsch,T., Merkl,R.,  
 Klockgether,J., Fritz,H.-J. and Tuemmler,B.  
 TITLE Direct Submission  
 JOURNAL Submitted (30-OCT-2001) Klinische Forschergruppe, Medizinische  
 Hochschule Hannover, Carl-Neuberg-Strasse 1, Hannover 30623,  
 Germany

FEATURES  
 source  
 1..128136  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="SG17M"  
 /db\_xref="taxon:287"  
 /map="between genes PA2836 and PA2815"  
 /clone="C"  
 /notes="hypervariable genome region"  
 misc\_feature  
 1..15865  
 /notes="corresponding genomic region to Pseudomonas  
 aeruginosa PAO1 from gene PA2836 to gene PA2820 with an  
 average nucleotide substitution rate of 0.27%; all  
 annotated genes are conserved and show only a few amino  
 acid substitutions"  
 tRNA  
 15712..15787  
 /product="tRNA-Glu"  
 15790..15865  
 /product="tRNA-Gly"  
 repeat\_region  
 15842..15865  
 /notes="att L site; flanks Pseudomonas aeruginosa SG17M  
 gene island PAGI-3(SG); 24 terminal nucleotides of the 3'  
 end of the Gly-tRNA gene"  
 /rpt\_type=direct  
 15866..119169  
 /notes="Pseudomonas aeruginosa SG17M gene island  
 PAGI-3(SG), integrated into the first Gly-tRNA gene"  
 15907..15934  
 /notes="AT-rich"  
 /rpt\_type=direct  
 16091..17500  
 /gene="int"  
 /notes="ORF SG1"  
 16091..17500  
 /genes="int"  
 /notes="bacteriophage P4 integrase subfamily; similar to  
 Xylella fastidiosa XF1718 (int) and Pseudomonas aeruginosa  
 C ORF C1"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="putative integrase"  
 /protein\_id="AAN62223.1"  
 /db\_xref="GI:24461654"  
 /translation="MALTDVAVRQARTGKDYTLDPFDGLSLAVSANGKSKWHFRYYM  
 AKQKMSLGTYPASVLRARALRDEARALAKGINKPKVDKHKLRVRLATENSFKA  
 VYLQMLHRLLEKRGQSTLSQIRIFDKDVLPLGTVTIFDIRSNLLDLVLAETQE  
 RGAFTAEKVRWTFROLFRFAMVKAEGVGNPASDLVVAAPKPPVAHNPFLRLHELQ  
 ELRLKRTGNITGOLGIRLLLTGVRTELELATPDQFDLERGLWIIIPPEVVKLOLQ  
 DGNRKGKQPDIPIPIVLSVQAEIVRYLLEQVAPQARHLLAHRGDLKXRISENIL  
 NAALRMVGLTGHTGIRGTJSTGNEIGYPKWVDAQLSHGLADPNKVSATYNHADIY  
 EPRKMQWADRLDLEQGDVKAASLHLAIHINGTQVQNESEGFLAIEGSAGTSVGN  
 SAQRNLGGATWTKQORM"  
 complement(17774..18835)  
 /gene="hemE"  
 /notes="ORF SG2"  
 complement(17774..18835)  
 /gene="hemB"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="putative uroporphyrinogen decarboxylase"  
 /protein\_id="AAN62224.1"  
 /db\_xref="GI:24461655"  
 /translation="MAELKNDRLRLRALREPVDVTPVMMRQAGRYLPEYRHRARAG  
 DFLSLCKNRELACEVTLPQLRYDLDAAILFSDILTIPDAMGLGLYFEEGEGPRFKKT

IRTEVEADLTVPDAASDLGYVLAAYKEIRSAAGRVPLIGFSGSPWTLATYVVEGGS  
 SKDFRHIKMMVATPEVLHALLDKLAQTVTGVLNEQIRSGVQAVQIFDTWCGVLSPGL  
 YLEFSLRYKMOIIVDGLIRSEGRVPVILFTKNGGOMLEAAEACADALGLDWTNID  
 DARRIGDRVALQNLDPGALYAPTKRAEVADILARFGSGSGHVFNLGHGHIQVVD  
 PEVVKVFDVAHELKSPYH"  
 19225..19680  
 /gene="ORF SG3"  
 19225..19680  
 /gene="ORF SG3"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAN62225.1"  
 /db\_xref="GI:24461656"  
 /translation="MNQREQSDNQVHLRLRWFLWLTWASIALAVLGALLPGLPTT  
 VFMLIAAWSASRCSQPLRWLEQHPFGDRLRNWQGVVIDRGSKWMAAGMLVSMAI  
 VLLSINQPLLVPIIATITGAVIWSRPEPSMNSQLMKQFDKNTY"  
 20010..20996  
 /gene="ORF SG4"  
 20010..20996  
 /gene="ORF SG4"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="hypothetical protein with methyl-accepting  
 domain of chemotaxis transducer"  
 /protein\_id="AAN62226.1"  
 /db\_xref="GI:24461657"  
 /translation="MLFGKVKSLSDETNYLSALMSGQIOIFRVDSLINQHPDLVEEC  
 LEDIQKWIQQSLAPSEPESTDELAANRSCLLNAQAEISYLRHELESSRSLIDE  
 ASHELSYRAEIIWDTKOTLTGECWELIVIDGIDNPANKLRKRFKSPRELIGVSTE  
 FELGDWSYVSIHPDDIKRVTKALDDYARIGNFDSYVVEYRMHKKIGDWTYREKG  
 RGVQPSQLRWIIGALRIDISDEKLAEMHARELENIQATYQISKVYGAIGIADQT  
 NMLALNAIEARAGVGFGFVADVKVLAAGRTAGRTREATQKIQEMLSDSKAF"  
 complement(21066..21818)  
 /gene="ORF SG5"  
 complement(21066..21818)  
 /gene="ORF SG5"  
 /note="originally longer protein; the integration of an  
 IS-element disrupted the CDS (ORFs SG8 and SG5)."  
 /codon\_start=1  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAN62227.1"  
 /db\_xref="GI:24461658"  
 /translation="MTELRRAINLDELFLVFPQKMEGTRRICGAELARWRHPSHGL  
 MSPDOFIYFAERSGLITRLTYWAIGAALRESYVWHSSGEAVPIAVNLSHDLRDPNLM  
 AHISSELTWGAHPDWIQFELTESCLMEDIANSMVLKMLRDLGFKLFDIDFGTYSS  
 LSYLRKLPVNYIKIDQSFVNNLGSDDSAIVRSTIDLAHNLGLEVAESGVESKIAMG  
 MLEANGCEVQGFPIKSKPIGSDPDFDWRLSFH"  
 22070..22477  
 /note="insertion sequence of the IS21 family; the  
 insertion sequence is flanked by inverted repeats and  
 contains a transposase and a NTP-binding protein; upon  
 insertion into ORF (SG5 and SG8) a sequence of 7 bp  
 (CCTAGT) was duplicated  
 /evidence=not\_experimental  
 22070..22477  
 /note="located on the gene island PAGI-3(SG)"  
 /insertion\_seq="IS21 family member"  
 22070..22076  
 /note="flanks insertion sequence"  
 /rpt\_type=direct  
 22077..22104  
 /note="terminal IR of insertion sequence"  
 /rpt\_type=inverted  
 22083..22137  
 /note="repeats L1 and L2 are inverted with respect to R1  
 and R2 found in the repeat region at the other end of the  
 insertion sequence"  
 /rpt\_type=direct

RESULT 15	
AP005021	
LOCUS	
DEFINITION	
ACCESSION	

```

VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Murata,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,
Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T.,
Kikuchi,H., Shiba,T., Sakaki,Y., Hattori,M. and Hattori,M.
Genome sequence of an industrial microorganism Streptomyces
avermitilis: deducing the ability of producing secondary
metabolites
Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
21477403
11572948
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
Ikeda,H., Ishikawa,J., Hanamoto,A., Shinose,M., Kikuchi,H.,
Shiba,T., Sakaki,Y., Hattori,M. and Omura,S.
Complete genome sequence and comparative analysis of the industrial
microorganism Streptomyces avermitilis
Nat. Biotechnol. 21 (5), 526-531 (2003)
22608306
12692562
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,
Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T.,
Kushida,N., Director-General of Biotechnology Center, Shiba,T.,
Sakaki,Y. and Hattori,M.
Direct Submission
Submitted (29-MAR-2002) Director-General of Biotechnology Center,
National Institute of Technology and Evaluation, Biotechnology
Center; 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
Tel:81-3-3481-1933, Fax:81-3-3481-8424
E-mail:bioente.go.jp, URL:http://www.bio.nite.go.jp/,
This work was done in collaboration with Haruo Ikeda(*1), Jun
Ishikawa(*2), Akiharu Hanamoto(*3), Chigusa Takahashi(*3), Mayumi
Shinose(*3), Hiroshi Horikawa(*4), Hidekazu Nakazawa(*4), Tomomi
Osonoe(*4), Norihiro Kushida(*4), Hisashi Kikuchi(*4), Tadayoshi
Shiba(*5), Yoshiyuki Sakaki(*6,*7), Masahira Hattori(*1,*7)
and Satoshi Omura(*1,*3).
Final finishing process and all annotation were done by H. Ikeda
and J. Ishikawa.
*1 Kitasato Institute for Life Sciences, Kitasato University
*2 National Institute of Infectious Diseases
*3 The Kitasato Institute
*4 National Institute of Technology and Evaluation
*5 School of Science, Kitasato University
*6 Institute of Medical Science, University of Tokyo
*7 RIKEN, Genomic Sciences Center
Following url is also available.
http://avermitilis.lis.kitasato-u.ac.jp.
Location/Qualifiers
1. 300550
/organism="Streptomyces avermitilis MA-4680"
/mol_type="genomic DNA"
/strain="MA-4680"
/db_xref="taxon:227882"
/notes="This strain is also named as strain: ATCC 31267,
NCIMB 12804 or NRRL 8165.-synonym: Streptomyces
avermectinius"
complement(1380..1811)
/notes="SAV1"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC67710.1"
/db_xref="GI:29603637"
/translation="MTAEWYLVEDTRETKEADGVELRLHWKLAATOHIAQDQEOA
AAAAADALNMGVLARHARPGDEPARHAFITQGANLVLRLQRRECHIRVTIARL
MHTQESKEAPPKFSFKLSALDGPQPEPAGRPWPGSET"
2466..2699
/notes="SAV2"
CDS
FEATURES
source
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC67711.1"
/db_xref="GI:29603638"
/translation="MPPQAGRGTIIVSATSGKTSMAAATLNCFFPEGRILVTVPTLIDL
LAQTAQAWRAVGHSHSPMIAVCSLENDPVLNERT"
2765..3310
/notes="SAV3"
/codon_start=1
/transl_table=11
/product="putative endonuclease VII"
/protein_id="BAC67712.1"
/db_xref="GI:29603639"
/translation="MDWNPDDDIFFCGCGDDTDPDRVPRQDKALCVRCDRVEROV
RRIYIVPRRNIMRFQDVCALCOEGPTDCHPDVAFSWHIDHHRCCPPGSGCGR
VGLICLPCNATRLPAYERLNPVLRSDSPFNLYLNSPPARHPPEARPTAKDHAGPRDAS
SYLIDAFPTAADHPGEGNALSS"
complement(3433..3939)
/notes="SAV4"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC67713.1"
/db_xref="GI:29603640"
/translation="MALTPGGTGVTVQWODROAIGDMHERRVAAALRARGTVPQCSOG
TYPYAVREALRTRTSALRHFPLIAARGADLITIDAKDMPSTDRYAVASLADTVTGA
LFFTAHAAPTPLYVFGDLKVLTPAEVHYHTAHLRHSRGAPHLVTEQAHCFFDDVFG
SAGAAAAA"
3994..4932
/notes="SAV5"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC67714.1"
/db_xref="GI:29603641"
/translation="MWMLMAAYVDRFRPTLPWGPVTPPELMPRLGARADGEWIVMT
POVRSRSHVPVPEDFYLRFEVMEVDPEDLDAVALMGAYHGLGSGINTSGMVDVVER
IKELTEREHPAPFALHGEIATLFWREAQAATTWLARREGGLDALIEPVESEELA
QWQANADLEAWPRDLHLRSLSEIRISNLVSLNAALKFSPSIGLGDYPTIL
AVAFQLYNHLAEDATIRECANETCRHFPVRQGRGAAAYGQNKTSIGIKYCTRCARAQA
QREHRRRRKQTTTLQPPAPGQSHDTSEPTAEGR"
4936..7563
/notes="SAV6"
/codon_start=1
/transl_table=11
/product="putative helicase"
/protein_id="BAC67715.1"
/db_xref="GI:29603642"
/translation="MISLREHOVEANARIRAWAGFTRSPVPAQGLRGTVSATSGSK
TITAAWAARECFRGRILVWPTDLLVQTAQMBRVGHNGPMVAACSLKEDVLEQL
GVRTTNPIQLAWAGHGCVVVFATYASLVREDDEDVTGRAKVGPLPAAALAGGRL
YQOTWDGFDLAVDEAHSITDGLRPWAAIHONSRIPADFRILYLTATPRILASPRQK
GADGRELIATWASDPDGPYGLSEAVERGILAGFEIDVLEIRDPDPALGES
EEAQGRRLALQTLALLHAAANRLTVMTFQRVEEAFAQTQPTQARLYEAEVS
AEALVDAGALPESSIGAEFYEACRHVPDPRVMAALCGDHLVAERREVLQFADGL
DAGNKRVRAPFLASVYVLGEGVDIVGERGVEAICFADTFGSOVELVQNLGRALRNPD
GINKTARIIVPVLPQENPTDMVASFAPLVTVLQGLRSRSEVLQALSRALTSG
QNHVVKDDEDR11GTTTTEGGGQHESEGAVESALLHFTSPRDTATTIAFLRIVYR
PESLVWLEGYQALLRWKKNHTGLYAVPYDTEAGVTKAPLGRWVHQDRTYRAG
ELDPHRTLLDAGMWPEGDDEANWLAALRSFRAHGLAPRDAVWGDADSELVP
VGERMANLRKDKLGNPORAATRAATLAAIDPWNCPWPLQWRHYRVLADLATDEP
HSRLPDIOGVQFEGDDLQKWLQRORSWAEISEEQORLTALGVTTPASPTTPSAK
GGGKAAAFORGLAALAAQVIREGAKHVPRGHVEAVIDGQEHQHLGWIINSKTRR
DKLTHDORTAALAAALGVEWA"
7554..9626
/notes="SAV7"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="BAC67716.1"
/db_xref="GI:29603643"
/translation="MGLAAGSASAVGVLLSDHPSEAGACCRQVCFQGAHASRWLSSVA

```

CWLUSSRWLFVRSSTKSRVCPVSRASSCCSPWRFREHRRTPPDAAVAPWAAQOPPY  
 LAPHQSLPELLELVYGLQOVDWELRIFFPAQVRVVRDLAETGLIDGIRDQ  
 QMCPRSTVAVLRMLGRVTRAVRGAHEATYGTAPQDDVLDRALQSRSTPAGIRPK  
 TDLRTTROPMLGLRLTWTVOORPGADEFARTLGRVELASRALAORPGTDPAFLRY  
 DVTAVDAIETALKRGEPPAGMYRMSIASHFENLDYGRHSNAADDLSAAFRDPA  
 LHRPEQANEDETKALPEPVIQOLDLHLLGLGRAGORTLAPDDQWYRTLYI  
 LRDTGQPLSVLSLPDCLRETRNDQISLIWNHKKARRHRRLEPITTSQAQIETWOE  
 RREGQASLPSTADYLFPLTHLATPRLYLSYLGELTRDWDVDSIPLHGEGETDQG  
 NPLFPDRSLIYAYAFRHSYAQRHADAGTPLDVLRLMDHKSVSITQRYTYVSLKRRQD  
 AVTKLAHVHDHGHPPSPSETAYELRSVAVPGGCTEPTSPNVKAAGSSCPTRFOCAGC  
 GFYRDPDLYPAISQHINELRADRETAQADAAEFVITAFQAITYAEQVTDNRQREL  
 AALPVAER"

## CDS

complement (9886..10635)

/note="SAV8"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAC67717.1"  
 /db\_xref="GI:29603644"  
 /translation="MSSISSALSRLPAPRAENAGDLDDGGWAAWLRLHGLDPAWRTSEW  
 RQDCWLFVGSVHEPRSSVALRCQTEACTVVSFANIFCPFKKEOKRSLPLPAEAPARAF  
 VVNRNVAFGVGPVPCFTKDGQRCVRPRHCKELCATHYTKWTKHSTRKTAGRWEDTA  
 VPYADTPACAPACPLGLXGRGLCRHHAORWMAFLIRLSLAVRISCLASSSSVRS  
 VAVLVRAGEGVVGCWRCCAAAMCWRMPSA"  
 complement (10632..10805)

## CDS

/note="SAV9"

/codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAC67718.1"  
 /db\_xref="GI:29603645"  
 /translation="MLRHSAAATRLRDGVDRDVRVQRLIGHASPLSMERYHRVDAEAR  
 AAVRVSGLKERR"  
 complement (11101..13497)

## CDS

/note="SAV10"

/codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAC67719.1"  
 /db\_xref="GI:29603646"  
 /translation="MLRSLACWPSRATDPSPVSELVPRPYAGTTTPOWQRLRLLEWAL  
 TPGILDAVLDELIESHADEARGPIAVNSDFWLSLYGLAETAPAPARLVGLVLRHLA  
 RARADSDGDPFESGHLSHTSOQADTVLSRTAEAPFAYVSQVLPFVLDVATAGSTGT  
 DAYAPSGRWARHLVGGHVSQAALTALDTRLSLAANS PAAADALQOLTPSPVQELR  
 FLACRVVYVNAQDEATLWLNDRNLRLGLWSLAWASRELIEATTPHCSDETLERL  
 TAMLIDYVPAWEHRRQSHAWCWSOYELLASICPSRRSDAGCRRLEAWEKPEPGOV  
 PSPPTIQTGVGSPISDHAARHMTNQWHRALNMVAKPQARERFPWPGGVHLEAQL  
 GSRAEQPERFTDFAFTLGDPAATYLCAIVGAVTPHLDDRWEOALHQLHILGPAA  
 APTICRALQSAQNFTAAASLPALSDYTDHPHEQDIRDADAEGTRTDLTLTAGMNATRG  
 QAALTVAALLFHGSEHLHALTPLVRLANDSVLAVRVCAQAVALAMKHPQIALDTT

## Alignment Scores:

Pred. No.: 4,4e-39 Length: 300550  
 Score: 554.00 Matches: 146  
 Percent Similarity: 49.67% Conservative: 79  
 Best Local Similarity: 32.23% Mismatches: 219  
 Query Match: 23.32% Indels: 10  
 DB: 1 Gaps: 6

US-10-098-602A-2 (1-454) x AP005021 (1-300550)

Qy 3 IleThrTyrAspGluLeuAsnAsnLeuIleAaGlyLysIleAspThrValValLeu 22  
 Db 227871 GTGACACTCGAAGAGCTCGGCCACTCTGTTCAAGCGCGTGAGATCGACACGGTATGGTC 227930  
 Qy 23 AlaCysValAspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHisPheLeu 42  
 Db 227931 GCCGTCGCGACCTGAGGGCGCGCTGATGGCAAGAACCTCGACGCCACCACTTCTGT 227990  
 Qy 43 GlyLeuAspGlnLysLysIleSerIleSerThrPheValTyrAlaValThrIleGluGly 62  
 Db 227991 ACCAAGCTGCCCGCGGACCGGAGATGTGGCCCTACATCTCTGGCGGACAGACGTGGACATG 228050

Qy 63 IleAlaGlyGlyGlyTyrGluIleSerSerValAspThrGlyTyrSerAspCysHisLeu 82  
 Db 228051 AACCGCTGGACGGTTCGAACCTCGCGGTGGCACACCGGATACGGAGACCTGGCTGTC 228110  
 Qy 83 CysAlaAspLeuAsnSerLeuHisLeuLeuProTrpSerGluGlyAlaValLeuAlaIle 102  
 Db 228111 GTCCCCACCTTGGGCACCATCCGCGCGCTGAACCTACCTGCGCAGGACCGTCTCTGTACAC 228170  
 Qy 103 SerAsnProHisAsnPheValThrSerGluProLeuPheCysSerProArgValIleLeu 122  
 Db 228171 GGCACCGCGTACAC---CCGACGCGCGCCCGCTCGCGCTGCACACACGGAGATCTCTC 228227  
 Qy 123 MetGlnGlnIleGluArgLeuAlaAsnLeuLysLeuLysGlyLeuPheAlaSerGluLeu 142  
 Db 228228 CGCAGGCAACTCGCGGCGCTCGCGCTCCACGCGCTGCAGGCACGCGTCGCGCATCGAGAGC 228287  
 Qy 143 GluPheAsnLeuPheAsnGluThrTyrLysSerAlaSerGlnLysHisTrpLysAsnLeu 162  
 Db 228288 GAATTCTGCTCTTCAAAAGGCACCGACGCTGCGCGCATCCGGGTTCAGGAACCCG 228347  
 Qy 163 LysThrAlaGlnProHisGlnTrpMetAsnIleSerAlaSerSerGlyIleGluThr 182  
 Db 228348 ACACCGCTGTCCTCCGACCAACCTCGACTACGCGCTGCACCGCCCGCCCTGAACCGG 228407  
 Qy 183 PheMetArgSerValArgAsnLysLeuGluAlaGlyIleLeuMetGluAlaThrHis 202  
 Db 228408 TTCTTCACGACCTGCACAAACGCGCTGCACGAGCAGCGGCACCAATCGAGGCCATCAA 228467  
 Qy 203 ProGluPheLeuProSerGlnHisGluLeuAsnPheValProAlaAsp---ProLeuThr 221  
 Db 228468 ACGGAAGGAGCCCGCAGGCACAGTCGAAAGTGGC-CTGGCCCTTACGGGACACCCCGCTGGGC 228526  
 Qy 222 MetAlaAspArgHisIleAlaLysHisGlyValArgGluMetAlaGluGlnSerGly 241  
 Db 228527 GCATGCGACACTACACCTCCACAAACACGCGCTCGTCATCTCGCCGCGCAACACCGC 228586  
 Qy 242 MetValAlaThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIle 261  
 Db 228587 CTCACACCCAGTTTCATGGCC---GCCCGCAGACCGCGCAGCATCGCGCTCCACCTG 228643  
 Qy 262 HisMetSerLeu---GlnAspAlaGluThrGluLysAsnAlaPheTyrAspGluAsnAsp 280  
 Db 228644 CACATCTCCCTGTGGCGCGACGACGAG------CGGCGCTTTCGCCATCGCTCCA 228691  
 Qy 281 GluTyrGlyMetSerThrLeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyrValPro 300  
 Db 228692 CGCGAGAGCTGCCCGAGCTACTCCACGACTCCATTCGCGGCTGTATCGACGACTGCGG 228751  
 Qy 301 GluAlaThrTyrPhePheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThr 320  
 Db 228752 CAGCTGGTTCGCTGTAGCGACCCACCCCACTCTCTACAAACGATATGTCCGCCACTCC 228811  
 Qy 321 PheAlaProThrLysCysCysTyrTrpAlaIleAspAsnArgThrSerAlaPheArgLeuCys 340  
 Db 228812 TTGCGCCCAACCAACTACACCTGGGACGCGCAACACCGCCTCGCGCATCTCGCATCACA 228871  
 Qy 341 AsnSerLysSerGluGlyIleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsnPro 360  
 Db 228872 GGGCACACGCC---TGCACACACTTGGAAAGTACGCTCGCGGCGCGGACGCCCAACCCC 228928  
 Qy 361 TyrLeuAlaPheSerAlaIleAlaAlaGlyIleSerGlyIleGluGluLysLeuGlu 380  
 Db 228929 TACTCGCACTCGCGGACCATCGCGCCATCGTCTACGCGCTCAACACCATCCCAAG 228988  
 Qy 381 LeuProProAlaSerGlyAsnValTyrAsnAspLysGluLeuProGluPheProAsn 400  
 Db 228989 CTCGCCGAGCCCTGCACCGCGACGCTACGCGCTCCGACCGCTCCCGCTCCCGCCACGC 229048  
 Qy 401 SerLeuGlnAsnAlaThrHisLeuLeuLysGluSerLysMetLeuAsnLysThrPheGly 420  
 Db 229049 AACCTCGAAGAGCCCTCACCGACTTCCAAACAAAGAGACGATCGCATCGCGCGCATTCAGC 229108  
 Qy 421 GluLysLeuIleLeuHisTyrValAsnAlaAlaAsnValGluIleAsnGluPheSerLys 440

```

Db 229109 ACACCCGTAGTCAGCACTAGCGCCGCGCTGCGGAGGTTGAGATCGAGGCACAGCACGGC 229168
      ::::: ||||| ||||| ||||| ||||| |||||
Qy 441 GlnValThrAspTrpGluLeuAenGlnGlyPheAenArg 453
      ||||| ||||| :: ||||| ||||| |||||
Db 229169 CAGGTCACAGACATCGACCGCAAGCGGGGTTCTCCCGT 229207
      ::||| ||||| ::||| ||||| |||||

```

Search completed: December 18, 2003, 00:54:26  
 Job time : 6869 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 22:42:10 ; Search time 19 Seconds  
(without alignments)

2297.925 Million cell updates/sec

Title: US-10-098-602A-2

Perfect score: 2376

Sequence: 1 MITYDELNNLRNGKIDTV.....INEFSQVTDWELNQGFNRY 454

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 76.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	942.5	39.7	454	2 AF2839	glutamine syntheta
2	942.5	39.7	454	2 H97616	probable glnA4 pro
3	927	39.0	434	2 A13578	glutamate-ammonia
4	799.5	33.6	457	2 F70885	probable glnA4 pro
5	798.5	33.6	462	2 T36848	glutamine syntheta
6	510.5	21.5	455	2 C84133	glutamine syntheta
7	480	20.2	439	2 B72313	conserved hypothet
8	469	19.7	413	2 C83225	glutamate-ammonia
9	466	19.6	443	1 AJCLOA	glutamate-ammonia
10	465	19.6	476	2 AB3575	glutamine synthase
11	464.5	19.5	449	2 G75517	glutamate-ammonia
12	464	19.5	446	2 H86899	glutamate-ammonia
13	459	19.3	458	2 B83609	probable glutamine
14	458	19.3	458	2 A83391	probable glutamine
15	450	18.9	452	2 D83609	probable glutamine
16	443	18.6	446	2 A43995	glutamate-ammonia
17	442.5	18.6	444	1 AJSBSQS	glutamate-ammonia
18	434.5	18.3	429	2 D90428	hypothetical prote
19	432	18.2	448	2 C95058	glutamine syntheta
20	432	18.2	448	2 D97927	glutamate-ammonia
21	430	18.1	449	2 H83944	glutamine syntheta
22	424.5	17.9	446	2 T51803	glutamate-ammonia
23	423	17.8	446	2 H89905	glutamate-ammonia
24	421	17.7	464	2 AE2650	glutamine syntheta
25	421	17.7	479	2 D97432	probable glutamine
26	412	17.3	445	2 A48947	glutamate-ammonia
27	411.5	17.3	444	1 AJSBSQU	glutamate-ammonia
28	411	17.3	444	2 AH1599	glutamine syntheta
29	410	17.3	444	2 AC1237	glutamine syntheta

#### ALIGNMENTS

##### RESULT 1

AF2839

glutamine synthetase glnA [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
C:Accession: AF2839

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell

i Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,

ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; MUID:21608550; PMID:11743193

A:Accession: AF2839

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-454 <KUR>

A:Cross-references: GB:AE008688; PIDN:AAL43132.1; PID:gi7740606; GSPDB:GN00186

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: glnA

A:Map position: circular chromosome

C:Superfamily: glutamate-ammonia ligase

Query Match 39.7%; Score 942.5; DB 2; Length 454;

Best Local Similarity 43.3%; Pred. No. 5.2e-69;

Matches 195; Conservative 89; Mismatches 165; Indels 1; Gaps 1;

QY 4 TYDELNNLRNGKIDTVVLACVDMOGRMLGKRLTGRHFLGLDQOKISISTFFVAVTIEGI 63

Db 5 TFDALKMDVAEGRITDTVLACLVDMOGRMLGKRLTGRHFLGLDQOKISISTFFVAVTIEGI 64

QY 64 AGCGYEIISVDYGYSDCHLCADNLHLPLWSEGAVALAISNPHNFVTSPLFCSPRVILM 123

Db 65 TVPGYKSSSWEGYGDYTLKPDLSLTKRVPWLEGTAFLVCLVDLHHHTEHVEVPSPEALLK 124

QY 124 QOIERLANLKLGLFASLEFNLFNETHYKASQKHWLKTAPHHQWNNISASSGIETFF 183

Db 125 KQVAREAMGLKAYMATELEFFLFDQTFDAARASGYKDLNLASGYNEDYHIQTTEEDV 184

QY 184 MESVANKLEEAGILMEATHPEPLPSOHEINFPADPLTMADRHITAKHGVREMAESQGV 243

Db 185 MEALRRGQAGIPVENSKEASPGQAEINVRVYAEALTWADRHAIKNATKRIAMSKGA 244

QY 244 ATFMAKLSSTALGNACHIMSIQDAETKNAFYDQNDYGMSTLARNWITAGLLKYVPEAT 303

Db 245 VTFLAKWYNNAAGSSSHIHQSLWLD-GKPAFLDKDGEHGMDSVMHYVAGLAHSDIT 303

QY 304 YFFASYSNKKQLPLTFAPTKCCWAINRTSAFLCNKSKSEGINVEIURIGADINPYIA 363

Db 304 YFLAPYINSYKRFMAGTFAPTAKIWSLDNRATGYRLCGAETKIRIECRVCGSDINPYIA 363

glutamine syntheta  
glutamate-ammonia  
glutamate-ammonia  
probable glutamine  
glutamate-ammonia  
glutamine syntheta  
probable gln2 - My  
glutamine synthase  
glutamine syntheta  
glutamate-ammonia  
glutamine syntheta  
glutamine syntheta  
glutamine syntheta  
probable glutamine  
conserved hypothet

401.5 16.9 460 2 AB2600  
31 400.5 16.9 446 2 T46736  
32 400.5 16.9 448 2 T45145  
33 400 16.8 452 2 C97382  
34 398.5 16.8 454 2 A64468  
35 395 16.6 442 2 F69076  
36 393.5 16.6 446 2 B70776  
37 391.5 16.5 448 2 A87113  
38 389.5 16.4 439 2 C75009  
39 387.5 16.3 454 2 T47127  
40 386.5 16.3 454 2 B84359  
41 386 16.2 491 2 E63368  
42 385.5 16.2 465 2 H87636  
43 381.5 16.1 439 2 A36911  
44 376 15.8 443 2 D71143  
45 375 15.8 454 2 D83449

QY 364 PSALIAAGISGIEEKLPPPPASGNVNDKELPEPPNSIQNATHLLKESKMLNKTFFGKL 423  
DB 364 MAALLAAGIDGLENKLELPAPVGGKQVREIPKTLRDATFLDGSKMLRQAFGDDV 423  
QY 424 ILHYVNAANVEINFSKQVDMELNOGNR 453  
DB 424 VDHVYHAAWQEEYDRRTDWEVARGFER 453

RESULT 2  
H97616  
probable glnA4 protein [imported] - Agrobacterium tumefaciens (strain C58, Cereon)  
C:Species: Agrobacterium tumefaciens  
C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002  
C:Accession: H97616  
R:Goodner, B.; Hinkley, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A:Reference number: A97359; MUID:21608551; PMID:11743194  
A:Accession: H97616  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-454 <KUR>  
A:Cross-references: GB:AE007869; PIDN:AAK87889.1; PID:gl5157281; GSPDB:GN00169  
C:Gene: AGR\_C3883  
A:Map position: circular chromosome  
C:Superfamily: glutamate-ammonia ligase

Query Match 39.7%; Score 942.5; DB 2; Length 454;  
Best Local Similarity 43.3%; Pred. No. 5.2e-69; Indels 1; Gaps 1;  
Matches 195; Conservative 89; Mismatches 165;

QY 4 TYDELNNLRNGKIDTVVLACVDMQGLMGRKLTGRHFLGLDQKKISISTFFVAVTIEGI 63  
DB 5 TFDALKMDVAEGRIDTVLACLVDQMGLMGKRFQAEFFVESAFETHSCNVVLATDME 64

QY 64 AGGGEISVDTGYSDCHCADLNSHLPLWSEGAVLAISNPHNVTSEPLFCSPRVILM 123  
DB 65 TVPGYKSSWEXGYDYLKPOLSLTRKVPWLEGTALVCLVDLHHTHEEVPHSPRALLK 124

QY 124 QOIERLANLKLGLPASELEFNLFNETYKSAQKHWKMLKTAQPHQHMNISASSGIETFRMRSVRNKLKLEAGIL 183  
DB 125 KQVARELAWGLKAYNATELEFFLFOOTDAARASYKDLNLASGYNEDYHIFQITKEEDV 184

QY 184 MRSVRNKLKLEAGILMEATHPEPLPSQHELNFVPADPLTMADRHIIAKHGVRMAEQSGMV 243  
DB 185 MRALRKLQAGIPVENSKEASPCQAEINVRVYAEALTMADRHIIKNAKEIAWSKGKA 244

QY 244 ATFMAKLSTALGNACHIMSLQDAETKNAPFYQNDYGVNSTLARNWIGLLKYVPEAT 303  
DB 245 VTFLAKWYNNAAGSSSHIQSLWILD-GKPAFLDKDGHGMSDVNRHYVAGLLAHASDIT 303

QY 304 YFFASYINSYKELQPLTAPTCKCAINRNTSAFRLCNKSEGINVELRIGADINPVLAF 363  
DB 304 YFLAPYINSYKRFMAGTAPTATWNSLDNRATGYRLCGAETKGRICRGVGSGLNPVLA 363

QY 364 PSALIAAGISGIEEKLPPPPASGNVNDKELPEPPNSIQNATHLLKESKMLNKTFFGKL 423  
DB 364 MAALLAAGIDGLENKLELPAPVGGKQVREIPKTLRDATFLDGSKMLRQAFGDDV 423

QY 424 ILHYVNAANVEINFSKQVDMELNOGNR 453  
DB 424 VDHVYHAAWQEEYDRRTDWEVARGFER 453

RESULT 3  
A13578  
glutamate-ammonia ligase (EC 6.3.1.2) [imported] - Brucella melitensis (strain 16M)  
C:Species: Brucella melitensis  
C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 03-Jun-2002

C:Accession: A13578  
R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,  
S.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letesee,  
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis  
A:Reference number: AD3252; PMID:11756688  
A:Accession: A13578  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-434 <KUR>  
A:Cross-references: GB:AE008918; PIDN:AAL53796.1; PID:gl7984727; GSPDB:GN00191  
C:Experimental source: strain 16M  
C:Genetics:  
A:Gene: BMBI10554  
A:Map position: II  
C:Superfamily: glutamate-ammonia ligase  
C:Keywords: ligase

Query Match 39.0%; Score 927; DB 2; Length 434;  
Best Local Similarity 43.9%; Pred. No. 9e-68; Indels 6; Gaps 3;  
Matches 192; Conservative 78; Mismatches 161;

QY 21 VLAC-VDMQGLMGRKLTGRHFLGLDQKKISISTFFVAVTIEGIAGGGEISVDTGYSD 79  
DB 1 MLACIVDMQGLRIGKRFYQGFVESGYDETHGCGNLLADDIDMEFPVPGYEAAGWDKGYD 60

QY 80 CHLCADLNSHLPLWSEGAVLAISN--PHNVTSEPLFCSPRVILMQLKGL 137  
DB 61 FVMPKPDJLSTLRLAPWLEKTAIVLCLVDLHDH---HDDLSSHSPRAVLKQVORLHBERGYRAY 117

QY 138 FASELEFNLFNETYKSAQKHWKMLKTAQPHQHMNISASSGIETFRMRSVRNKLKLEAGIL 197  
DB 118 FASELEFYIDETYSARAKRWHEMETASPVQGVYIHLITREPPVLARMNHLADAGIP 177

QY 198 MEATHPEPLPSQHELNFVPADPLTMADRHIIAKHGVRMAEQSGMVATFMAKLSTALGN 257  
DB 178 VENSKEGSGPQQLNRYCKALEMDRHVIMKAMKEIAEAHGKCTFMKAYDYARAGS 237

QY 258 ACHIMSLQDAETKNAPFYQNDYGVNSTLARNWIGLLKYVPEATYFFASYINSYKELQ 317  
DB 238 SSHVNSIWSADGKBPLFPKAPYTWLTMRSWVAGQIKYATDYITFLAPYINSYKRFQ 297

QY 318 PLTPAPTCKCAINRNTSAFRLCNKSEGINVELRIGADINPVLAFSAIIAAGISGIEE 377  
DB 298 AGTFAPTCKMWSQDNRTAGFLCGEGTKGRICRGADINPVLAFALIAAGLKGVD 357

QY 378 KLEUPPPASGNVNDKELPEPPNSIQNATHLLKESKMLNKTFFGKLILHYVNAANVEIN 437  
DB 358 KLELDEPFGDAYSARLKEIPYTLREAAAALKGSFLKEAFGDDVNVNHYTHTAHWEQIE 417

QY 438 FSKQVTDWELNOGNRY 454  
DB 418 YDRVTDWELVYRGERY 434

RESULT 4  
F70885  
probable glnA4 protein - Mycobacterium tuberculosis (strain H37Rv)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: F70885  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;  
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.;  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.;  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: F70885  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-457 <COL>  
A:Cross-references: GB:AL008883; GB:AL123456; NID:G3261490; PIDN:CAA15522.1; PID:G261280





Query Match 19.6%; Score 466; DB 1; Length 443;  
Best Local Similarity 28.0%; Pred. No. 4.5e-30;  
Matches 130; Conservative 84; Mismatches 211; Indels 40; Gaps 8;

QY 4 TYDELNNLRNGKIDTVLACVDMQGRMLGKRITGRHFLGLDOKKISISIFVAVATIEGI 63  
DB :  
5 TKEDIINLVKENGVKIRLOFTDI FGTLKVAITDKQL-----EKALONECMFGSSIDGF 60  
QY 64 AGGYEISSVDTCYSCHLCADLNLSHLPLW--SEGAVLAISNPHNFVTSEPLFCSPRVI 121  
DB :  
61 -----VRIBESDNLRPNLDSFVIPWRPQQGVARLICDVYKPDGTPPEGDPRHV 111  
QY 122 LMQOIERLANLKLKGL-----FASELEFNLFNETYKSASQKHKNLKTAPHPHQWNISA 176  
DB :  
112 L-----KRANADAKELGYTMNVGPCEFFLFE-----TDENGRATTNTQDKAGYFDLP 160  
QY 177 SSGIETFMRSVRNKLEEAGILMEATHPEPLPSQHLELNFVPADPLTWADRHI IAKHGVRM 236  
DB :  
161 TDLGENARRDMTALAEMGEFEBASHHEVAGQNEIDFKYGDALTATTADNIMTFKLVVXSI 220  
QY 237 AEQSGMVATPMAKLSSTALGNACHIIMSLODAETERKNAPFYDONDEYGMSTLARNWTAGLL 296  
DB :  
221 AQRHGLHAGMPKPICFINGSGMHVANSU--PRDGKNAFVDENDKXGLSKVAVQFIAGLL 278  
QY 297 KYVPEATYFPASVINSYKRLQPTFAPTCCWAIDNRTSAPFLCNSKSEGINVELRIGGA 356  
DB :  
279 KNIKGMAAVTNPJLVNSYKSLPGVEAPVYLAWSCNRALTALIRVPAARGAGTRVELRCDDP 338  
QY 357 DLNPLYAFSAIIAAGISGTEKELELPPASGVNY-----NDKELPEPNSLQNATHLL 409  
DB :  
339 SSNPYLVLACLQGLDGLCKNNLQPPAEVANIFAMTEQERKENGINDLNPNNLYEAVNYM 398  
QY 410 KESKMNLKTFGEKLIILHYVNAANVEINFESKVQTVDELNQGFNRY 454  
DB :  
399 KENELAKCALGDHVYGNVYAGKAEDDYRTKVHDWELENLYNKY 443

RESULT 10  
AB3575  
glutamate-ammonia ligase [EC 6.3.1.2] [imported] - Brucella melitensis (strain 16M)  
C:Species: Brucella melitensis  
C>Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 03-Jun-2002  
C:Accession: AB3575  
R.; DelVecchio, V.G.; Kapratral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,  
.; Mazur, M.; Goldstein, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letesee  
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis  
A:Reference number: AD3252; PMID:11756688  
A:Accession: AB3575  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-476 <CUR>  
A:Cross-references: GB:AE008918; PIDN:AAL53765.1; PID:g17984693; GSPDB:GN00191  
A:Experimental source: strain 16M  
C:Genetics:  
A:Gene: BMEI10523  
A:Map position: II  
C:Superfamily: glutamate-ammonia ligase  
C:Keywords: ligase

Query Match 19.6%; Score 465; DB 2; Length 476;  
Best Local Similarity 28.7%; Pred. No. 6.1e-30;  
Matches 133; Conservative 71; Mismatches 208; Indels 52; Gaps 10;



A;Reference number: A82950; MUID:20437337; PMID:10984043

A;Accession: A83391

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-458 <STO>

A;Cross-references: GB:AE004630; GB:AE004091; NID:g9948041; PIDN:AG05428.1; GSPDB:GN001

A;Experimental source: strain PA01

C;Genetics:

A;Gene: PA2040

C;Superfamily: glutamate-ammonia ligase

Query Match 19.3%; Score 458; DB 2; Length 458;

Best Local Similarity 28.2%; Pred. No. 2.1e-29;

Matches 125; Conservative 85; Mismatches 211; Indels 22; Gaps 10;

Qy 16 KIDTVVLACVDMQGRMLKRLTGRHFLGLDQKKISISTFVAVTIG-----IAGGVEIS 71

Db 22 EVQFVDLLISDMGVGVRKRIENSLPKVPEKGINLPASLPALDITSTVESTGLGLDIG 81

Qy 72 SVDTGYSCHLCADL-NSLHLLPWSGAVLAISNPHNFVTSEPLFCSFVRVILMQQIERLA 130

Db 82 DAD-----RICYPITPGTSMPEWQKRPTAQLLMTMEHELECEPFADPREVLQVVARFT 135

Qy 131 NLKLGIFASELEFNLFN-ETYSASQKHWNKLTQAHPHQMWNISASSGIETFMRSVRN 189

Db 136 EMELTIVAFALEFYLIDQENVNGRPQPPRSPISGRP--OSVQVYSIDDLDEYVECLQD 193

Qy 190 KLEA---GILMEATHPEFLPSQHELNF-VPADPLTMADRHIIAKHGVREMAEQSGMVAT 245

Db 194 IIDGARAQIPADAIVAESAPAEQEVNLHVADPMKACDVAVLKRLIKNIAYDHEMDTT 253

Qy 246 FMAKLSLTALGNACHIHMSLODAETERKNAFYDQNDYGMSTLARNIAGLLKYVPEATYF 305

Db 254 FMAKYPQAGLGLHVHLSLD-KHGNNIPTSDPE--QNAALRAHIGGVLETLPASMAF 310

Qy 306 FASVINSYKRLQPLTFAPTCKCWAIDNRTSAFRLCNSKSEGINVELRIGGADLNPYLAFS 365

Db 311 LCPNVNSYRRFGSGFYVPNAPSGLDNRITVALRVPTGSPAVRLEHVRVAGADANPYLLA 370

Qy 366 AIIAAGISGIEEKLPPASGVNVDKELPEPPNSLQNAATHLLKESKMLNKTGKELIL 425

Db 371 SVLAGVHGLTNKVEPGAPTEGNSYEQLE-PSLPNNLRDALRELDSDSEILAKYIDPKYID 429

Qy 426 HYNAANVEINEFSKQVTDWELN 448

Db 430 IFVACKESLEEFYSISDLEYN 452

#### RESULT 15

D83609

probable glutamine synthetase PA0298 [imported] - Pseudomonas aeruginosa (strain PA01)

C;Species: Pseudomonas aeruginosa

C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 24-May-2001

C;Accession: D83609

R;Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A;Reference number: A82950; MUID:20437337; PMID:10984043

A;Accession: D83609

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-452 <STO>

A;Cross-references: GB:AE004467; GB:AE004091; NID:g9946133; PIDN:AG03687.1; GSPDB:GN001

A;Experimental source: strain PA01

C;Genetics:

A;Gene: PA0298

C;Superfamily: glutamate-ammonia ligase

Query Match

Best Local Similarity 18.9%; Score 450; DB 2; Length 452;

Matches 130; Conservative 80; Mismatches 222; Indels 26; Gaps 7;

Qy 1 MTITYDELNNLIRNGKIDTVVLACVDMQGRMLKRLTGRHFLGLDQKKISISTFVAVTII 60

Db 1 MTTKLDQLTSWLKERKITEVECLISDLTGIARGKISPTNKFIAEKGMRLPESVLLQTVTG 60

Qy 61 EGIAGGVEISSVDYTGSDCHLCADLNSLHLLPWS-EGAVLAISNPHNFVTSEPLFCSPR 119

Db 61 DYVEDDIY-YDLLDPADIDMVCPRDENAVFLVPAIEPTAMVIHDTFDKL-GNPIELSPR 118

Qy 120 VILMQQIERLANLKLGLFASLEFNLFNET-----YKSASQKHWNKLTQAQPH 168

Db 119 NILKRVLKNYADKGNRPVIVAPENEFYLTKESSDDPDYPLQAPVGRSGRQ-----ETG 169

Qy 169 HQMWNISASSGIETFMRSVRNKLKEEAGILMEATHPEFLPSQHELNFVPADPLTMADRHII 228

Db 170 RQSFSDAANEFDPLEDMYDWCEAQGLDLTLIHEBGTQAMEINFRHGDALDLADQILV 229

Qy 229 AKHGVREMAEQSGMVATFMAKLSLTALGNACHIHMSLODAETERKNAFYDQNDYGMSTLA 288

Db 230 FKRTMREAAALKHNVAATFMAKPMTGEPGSAMHLHQSIQVDTGKNIIF--SNADGTMSELF 287

Qy 289 RNMIAGLLKYVPEATYFFASVINSYKRLQPLTFAPTCKCWAIDNRTSAFRLCNSKSGIN 348

Db 288 LHHIGGLQKQFPEVLPFLFAPNVNSFRFLPDTAPVNVWEGEENRTVGLRVPDSSPENRR 347

Qy 349 VELRIGGADLNPYLAFSAIIAAGISGIEEKLPPASGVNVDKELPEPPNSLQNAATHL 408

Db 348 VENRLAGADANPYLAAASLLCGYIGWVEGIKPSAQVKGGRYERNL-RLPLTIEAALER 406

Qy 409 LKESKMLNKTGKELILHYNAANVEINEFSKQVTDWE 446

Db 407 MENCKPLEQYLGSKFISGYVAVKRAEHENFKRVISSWE 444

Search completed: December 17, 2003, 22:44:47

Job time : 21 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 22:40:45 ; Search time 38 Seconds  
(without alignments)

3083.050 Million cell updates/sec

Title: US-10-098-602A-2

Perfect score: 2376

Sequence: 1 MTTYTELNLNRNGKIDTV.....INBFSQVTDWELNQGNRY 454

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1015	42.7	465	16	Q98H15
2	969.5	40.8	455	16	Q8FVU3
3	963	40.5	454	16	Q92NI9
4	942.5	39.7	454	16	Q8UD13
5	927	39.0	434	16	Q8YCH5
6	799.5	33.6	457	16	Q33342
7	798.5	33.6	462	16	Q88070
8	614.5	25.9	459	2	Q93670
9	595	25.0	460	2	Q8GPY7
10	510.5	21.5	455	16	Q9K665
11	470.5	19.8	427	2	Q9AEL4
12	470.5	19.8	446	16	Q8NNH4
13	470.5	19.8	516	16	Q8NHL7
14	469	19.7	413	16	Q9HYP0
15	465	19.6	476	16	Q8YCK5
16	465	19.6	476	16	Q8FVQ7

17	464.5	19.5	449	16	Q9RX64
18	459	19.3	458	16	Q916J5
19	458	19.3	458	16	Q91275
20	453	19.1	445	16	Q8G4S3
21	453	19.1	461	16	Q8P897
22	451	19.0	448	16	Q8DVU9
23	450	18.9	452	16	Q916J3
24	450	18.9	478	16	Q92RY1
25	449.5	18.9	447	17	Q8PY99
26	449.5	18.9	483	17	Q976N5
27	449	18.9	448	17	Q9H149
28	449	18.9	461	16	Q8FJQ2
29	447.5	18.8	445	16	Q8EQP3
30	447.5	18.8	455	16	Q92MV2
31	446.5	18.8	491	17	Q8TYG1
32	445	18.7	448	16	Q8DXT3
33	445	18.7	451	16	Q8EHF4
34	444.5	18.7	447	17	Q8TID7
35	442	18.6	456	16	Q98EM0
36	441	18.6	448	16	Q99Y41
37	441	18.6	448	16	Q8NZG4
38	440	18.5	448	16	Q8E3F3
39	434.5	18.3	429	17	Q97VQ9
40	434	18.3	462	17	Q977Z8
41	432	18.2	448	16	Q97S86
42	432	18.2	448	16	Q8DQX6
43	430	18.1	449	16	Q9KAC9
44	430	18.1	478	16	Q988Z8
45	425.5	17.9	453	16	Q9RDS6

## ALIGNMENTS

### RESULT 1

Q98H15 ID Q98H15 PRELIMINARY; PRT; 465 AA.  
AC Q98H15;  
DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Glutamine synthetase;  
GN MLL3074;  
OS Rhizobium loti (Mesorhizobium loti).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Phyllobacteriaceae; Mesorhizobium.  
OX NCBI\_TaxID=381;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF303099;  
RX MEDLINE=21082930; PubMed=11214968;  
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,  
RA Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
RT Mesorhizobium loti";  
RL DNA Res. 7:331-338(2000).  
DR ENBL; AP003001; BAB50051.1; -;  
DR InterPro; IPR01637; GlnA\_adenyltn.  
DR InterPro; IPR001691; Gln\_synth.  
DR Pfam; PF00120; gln-synt\_1;  
DR Pfam; PF03951; gln-synt\_N; 1;  
DR ProDom; PD001057; GlnA\_adenyltn; 1.  
KW Complete proteome.  
SQ SEQUENCE 465 AA; 52610 MW; 99A46709D3D184FA CRC64;

Query Match 42.7%; Score 1015; DB 16; Length 465;  
Best Local Similarity 45.7%; Pred. No. 1e-75;  
Matches 206; Conservative 81; Mismatches 164; Indels 0; Gaps 0;

OY 4 TYDELNLNRNGKIDTVTLACVDMQGRMGKRLTGRHFLGLDQKTKISISTFVAVTIEGI 63

```

Db      15  SFQDLKKA VSSGEIDTVLACIVDMQGRLAGKFLAQYFVDSAHDETHGNCYLLAADIDWE 74
QY      64  AGGGYEISSVDGTGYSDBCHLCA DLNSHLPLPSEGA VLAISNPHNFVTSEPLFCSPRVILM 123
Db      75  PVPYGKAASWSKGYGDFVNMKPLATLRRTPWLKTA LVCVDLDDHDDLPSPRAILK 134
QY      124 QOIERLANLKLGLFASLEPFLNETYKSAQKHWNKLTAPQHHQWNNISASSGIEFF 183
Db      135  KQVKLSERGYIGYFASLEFLFNETYDSARKHWOGLDTASPIYDYGIGITTKERG 194
QY      184  MRSVRNKLKEEAGILMEATHPEFLPSQHLELNFVPADPLTMADRHIIIAKHGVREMAQSGMV 243
Db      195  MRLRLNEMEAAGPIPIENSGKGGWPGQGEINVRVYAEALDMDADRHVILKNGAKEIASSEGKA 254
QY      244  ATFMKALSTALGNACHIHMSLODAETEKNAFYDQNDYGVNSTLARNIAGLLKTVPEAT 303
Db      255  ISFMSKYNKYNGLAGNSSHINLSWSADGKTPFLFDKKADWTLSTLGGQWAAAGQLYAKEFT 314
QY      304  YFFASVINSYKRLQPLTEAPTCKCWAIDNRTSAPFLCNKSGSEGINVELRIGGADINPYLA 363
Db      315  WFLAPYINSYKRFQAGTFAPTKIMMSSEDRNTAGFLRCGEGTKGIRMECRIGGADINPYLA 374
QY      364  FSAIIAAGISGIEEKLPPPPASGNVYNDKELPEFPNSLQNA THLLKESKMLNKTFFGSKL 423
Db      375  FAALIAAGLAGIDEXLEKLQKPFVGDYQASRLPEIPKTLRDATE TLAKSKMLKQALGEDV 434
QY      424  ILHYVNAANVEINERFSKQVTDWELNQGFNRY 454
Db      435  LEHYVHTAKWQEFYDRITDWEHLRGPERY 465

```

## RESULT 2

```

Q8FVU3 ID Q8FVU3 PRELIMINARY; PRT; 455 AA.
AC Q8FVU3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Glutamine synthetase: family protein.
OS Brucella suis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29461;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=1330 / Biovar 1;
RX MEDLINE=22247741; PubMed=12271122;
RA Paulsen I.T., Seehadi R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
RA Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
RA Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.M.;
RT "The Brucella suis genome reveals fundamental similarities between
RT animal and plant pathogens and symbionts."
RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153 (2002).
DR EMBL; AB014569; AAN33914.1; -.
DR TIGR; BRA0732; -.
KW Complete proteome.
SQ SEQUENCE 455 AA; 51884 MW; ED78D55BA25C8E2B CRC64;

```

```

Query Match 40.8%; Score 969.5; DB 16; Length 455;
Best Local Similarity 43.6%; Pred. No. 6.1e-72;
Matches 198; Conservative 82; Mismatches 169; Indels 5; Gaps 2;
QY 3 ITYDELNNLRNGKIDTVVLACVDMQGRILMGKRLTGRHFLGLDQKKISISTEYVAVTLEG 62
Db 5 LTFDELKKA VADGRIDTVLACVDMQGRILMGKRLTGRHFLGLDQKKISISTEYVAVTLEG 64
QY 63 IAGGGYEISSVDGTGYSDBCHLCA DLNSHLPLPSEGA VLAISNPHNFVTSEPLFCSPRV 120

```

```

Db      65  BPVPGYEAAGWDKGYGDFVNMKPLDSTLR LAPWLEKTAI VLCDVLDHH---HDDLSHSPRA 121
QY      121  ILMOQIERLANLKLGLFASLEPFLNETYKSAQKHWNKLTAPQHHQWNNISASSG 180
Db      122  VLKQVQRLHGRGYRAVFASELEFYIFDETYKSARAKRHEMETASPVQDYVHLTTRE 181
QY      181  ETFMRSVRNKLKEEAGILMEATHPEFLPSQHLELNFVPADPLTMADRHIIIAKHGVREMAQSG 240
Db      182  EPLVLRAMNHLADAGIPVENSKGWPGQGEINVRVYAEALDMDADRHVILKNGAKEIAEAH 241
QY      241  GMVATFMKALSTALGNACHIHMSLODAETEKNAFYDQNDYGVNSTLARNIAGLLKTV 300
Db      242  KGCITFMKAYDARAGSSSHVHNSIWSADGKEPLFFDPKAPYTWTPLMRSVAVAGQIKYAT 301
QY      301  EATYFFASVINSYKRLQPLTEAPTCKCWAIDNRTSAPFLCNKSGSEGINVELRIGGADINP 360
Db      302  DYTFLAPYINSYKRFQAGTFAPTKIMMSQDNRTAGFLRCGEGTKGIRIECRIGGADINP 361
QY      361  YLAFSAIIAAGISGIEEKLPPPPASGNVYNDKELPEFPNSLQNA THLLKESKMLNKTFFG 420
Db      362  YLAFAALIAAGLKGVDKLELDEFPVGDAYSAVRLKEIPYTLREAAAALKGSFLKEAFG 421
QY      421  EKLILHYVNAANVEINERFSKQVTDWELNQGFNRY 454
Db      422  EDVNVNHYHTAHWEQIEYDRRTDWEHLRGPERY 455

```

## RESULT 3

```

Q92NI9 ID Q92NI9 PRELIMINARY; PRT; 454 AA.
AC Q92NI9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein R02213.
GN R02213 OR SMC01594.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramepeger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882 (2001).
DR EMBL; AL591790; CAC46792.1; -.
DR InterPro; IPR001637; GlnA_adenyltn.
DR InterPro; IPR001691; Gln synth.
DR Pfam; PF00120; gln-synt_1.
DR Pfam; PF03951; gln-synt_N; 1.
DR ProDom; PD001057; GlnA_adenyltn; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 454 AA; 51112 MW; B4265B789114D431 CRC64;

```

```

Query Match 40.5%; Score 963; DB 16; Length 454;
Best Local Similarity 43.7%; Pred. No. 2.1e-71;
Matches 198; Conservative 87; Mismatches 169; Indels 0; Gaps 0;
QY 1 MTITYDELNNLRNGKIDTVVLACVDMQGRILMGKRLTGRHFLGLDQKKISISTEYVAVTI 60
Db 1 MSYFEELEKEDVAAGRIDTVLACQVDMQGRILMGKRFHAEYFVESAWKETHSCNYLLATDM 60
QY 61 EGIAGGGYEISSVDGTGYSDBCHLCA DLNSHLPLPSEGA VLAISNPHNFVTSEPLFCSPRV 120
Db 61 EMETVPGYKATSWKGYGDTYMKPDLSTLRIPMLLEGTA LVLCMDLHDHDTAEVPHSPRA 120

```

```
Qy 121 ILMOQIERLANLKLGLFASLEENLNFTYKSAQKHWNKLTAPQPHOMNITSASSGI 180
Db 121 ILKQVAREANGFKAYMASSELEFFLDQSYYDARLSGYDLQALSGYNYDYHIFQTTKE 180
Qy 181 ETFMRSVRNKLLEAGILMEATHPEFLPSQHELFNVPDPLTMADRHIIAKHGVRMAEQS 240
Db 181 EDVNRALRNGLOAGIPVENSKGASAGQEEINRYADALTMADRHAIKNGCKEIAWQR 240
Qy 241 GMVATFWAKLSSTALGNACHTHMSLQDAETKNAFYDNDYGNSTLARNWIAGLLKVP 300
Db 241 GKAITFLAKWNSAAGSSSHIHQSLWSKOGTFLFFDKNGQYGNSELMRHRVVAQLAHAS 300
Qy 301 EATVFFASYINSYKRLQPLTFAPTKCCWAIDNRTSAFRLCNSKSEGINVELRIGADLNP 360
Db 301 EVTYFLAPYINSYKRFNAGTAPTKAIWSKDNRTAGYRLCGESKAIKRIECRVGSDLNP 360
Qy 361 YLAFSAITAGISGIEKLELPPASGNVNDKELPEPPNSLONATHLLKESKMLNKTFF 420
Db 361 YLAFALTAAGIAGIENKMLEAPFVGDAYGQKVEIREIPHTLRGALSGSKMLRAAFG 420
Qy 421 EKLILHYVNAANVEINFSKQVTDWELNQGFR 453
Db 421 EEEVDHYVHAARWEQYEDRRVTDWEVARGFER 453

RESULT 4
Q8UDI3
ID Q8UDI3 PRELIMINARY; PRT; 454 AA.
AC Q8UDI3;
DT 01-JUN-2002 (TrEMBLrel. 21; Created)
DT 01-JUN-2002 (TrEMBLrel. 21; Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23; Last annotation update)
DE Glutamine synthetase.
GN GLNA OR ATU2142 OR AGR C 3883.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Neeter E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RT Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmlel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RT Science 294:2323-2328(2001).
RN [3]
RX MEDLINE=21608552; PubMed=11743195;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmlel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RT Science 294:2323-2328(2001).
RN [4]
RX MEDLINE=21608553; PubMed=11743196;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmlel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RT Science 294:2323-2328(2001).
```

```
KW Complete proteome.
SQ SEQUENCE 454 AA; 50703 MW; 559465BB5613E37E CRC64;

Query Match 39.7%; Score 942.5; DB 16; Length 454;
Best Local Similarity 43.3%; Pred. No. 1.1e-69;
Matches 195; Conservative 89; Mismatches 165; Indels 1; Gaps 1;

Qy 4 TYDELNLLIRNGKIDTVVLACVDMQGRMLGKRLTGRHFLGLDQKKISISTFVYAVTIEGI 63
Db 5 TFDALKMDVAEGRIDTVLACVDMQGRMLGKRLTGRHFLGLDQKKISISTFVYAVTIEGI 64
Qy 64 AGGGEYISSVDVTGYSDCHLCLADLNSHLPLWSEGAVLAINPHNFVTSEPLFCSPRVILM 123
Db 65 TVPGYKSSSEWEGYDYLKLPDLSTLRKVPWLEGTALVCLDVLDDHHTHEEVPHSPRALIK 124
Qy 124 QOTERLANLKLGLFASLEENLNFTYKSAQKHWNKLTAPQPHOMNITSASSGIETFF 183
Db 125 QVARELMEAGILMEATHPEFLPSQHELFNVPDPLTMADRHIIAKHGVRMAEQSGMV 184
Qy 184 MRSVRNKLLEAGILMEATHPEFLPSQHELFNVPDPLTMADRHIIAKHGVRMAEQSGMV 243
Db 185 MRALRKGLQAGIPVENSKGASAGQEEINRYADALTMADRHAIKNGCKEIAWKGKA 244
Qy 244 ATFWAKLSSTALGNACHTHMSLQDAETKNAFYDNDYGNSTLARNWIAGLLKVPBEAT 303
Db 245 VTFLAKWNSAAGSSSHIHQSLWSLD-GKPAFLDKGEGHGMDSVMRHRVVAQLAHASDIT 303
Qy 304 YEFASYINSYKRLQPLTFAPTKCCWAIDNRTSAFRLCNSKSEGINVELRIGADLNPVLA 363
Db 304 YFLAPYINSYKRFNAGTAPTKAIWSDNRTAGYRLCGESKAIKRIECRVGSDLNPVLA 363
Qy 364 FSAITAGISGIEKLELPPASGNVNDKELPEPPNSLONATHLLKESKMLNKTFFGKEL 423
Db 364 MALLAAGIDGIEKLELPPASGNVNDKELPEPPNSLONATHLLKESKMLNKTFFGKEL 423
Qy 424 ILHYVNAANVEINFSKQVTDWELNQGFR 453
Db 424 VDHYVHAARWEQYEDRRVTDWEVARGFER 453

RESULT 5
Q8YCH5
ID Q8YCH5 PRELIMINARY; PRT; 434 AA.
AC Q8YCH5;
DT 01-MAR-2002 (TrEMBLrel. 20; Created)
DT 01-MAR-2002 (TrEMBLrel. 20; Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22; Last annotation update)
DE Glutamine synthetase (EC:6.3.1.2).
GN BMEI10554.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=16M / ATCC 23456 / Biotype 1;
RX MEDLINE=2002109; PubMed=11756688;
RA DelVecchio V.G., Kapral V., Redkar R.J., Patra G., Mujar C., Los T.,
RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
RA Jablonksi L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltzman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Haselkorn R., Kyrtides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis."
RT Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
RN [2]
RX ENBL; AE009691; AAL53796.1;
DR InterPro; IPR001637; GlnA adenyln.
DR InterPro; IPR001691; GlnA synth.
DR Pfam; PF00120; gln-synt; 1.
DR Pfam; PF03951; gln-synt; 1.
DR ProDom; PD001057; GlnA adenyln; 1.
KW Ligase; Complete proteome.
SQ SEQUENCE 434 AA; 49582 MW; 1114E54225746C7A CRC64;
```



```

RA Oliver K., Harris D.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN (3)
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
RN Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RL the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RN Mol. Microbiol. 21:77-96(1996).
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RN Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RN Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,
RN Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RL coelicolor A3(2).";
RL Nature 417:141-147(2002).
CC -1- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + NH(3) = ADP + PHOSPHATE +
CC L-GLUTAMINE.
CC -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
DR EMBL; AL939109; CAA20824.1; -.
DR HSSP; P06201; ILGR.
DR InterPro; IPR001637; GlnA_adenyltn.
DR InterPro; IPR001691; Gln_synth.
DR Pfam; PF00120; gln-synt_1.
DR ProDom; PD001057; GlnA_adenyltn; 1.
DR Ligase; Complete proteome.
KW Ligase.
SQ SEQUENCE 462 AA; 50818 MW; 595825AA19EBE700 CRC64;

Query Match 33.6%; Score 798.5; DB 16; Length 462;
Best Local Similarity 38.9%; Pred. No. 9.9e-58;
Matches 176; Conservative 84; Mismatches 177; Indels 15; Gaps 8;

Qy 6 DELNNLRNGKIDITVVLACVDMQGLMKRLTGRHFLGDLQKISIST----FYAVTIE 61
Db 21 EELHALVAAGDIDITVVLAFDPDMQGLRQKRFPAARFL--DEVLEHGTGECNYLLAVDAD 77
Qy 62 GIAGGGEYISSVDTGYSDCHCADLNSLHLLPWSGAVLAISNPHNFTVTSPLFCSPRV 121
Db 78 MNTVDGYAMSSDRGYGDFANRADPATLRLPWNEGTAMAVAD--LAWEDGSPVLAAPRQI 136
Qy 122 LMQOIERLANLKLGLFASLEFNLFNETYKSAQKHWNKLTAPQHHQWNISASSGIE 181
Db 137 LRRQLERLHGYTAQVTELEFIVFRDYTEHAWDANYRGUTPANQNVDSVILGTGRVE 196
Qy 182 TFMRSVRNKLREAGILMEATHPEFLPSQHELNFPADPLTMADRHIIAKHGVREMAEQS 241
Db 197 PLLRRIRNEMAGLTVESAKGECNPGQHEATFYDEALVTCQDHVYKTKAKIAQEG 256
Qy 242 MVATFMKLSLTALGNACHHMSIQDAETKNAFYQNDYQNDYQNDYQNDYQNDYQNDY 301
Db 257 MSLTFMAYKNELE--GNSCHILSLADAD--GRNMAEGG--GMSDVMRHFVLAGQVLA 311
Qy 302 ATYFFASVINSYKRLQPLTFAPTCCWADNRITSAFLCNKSKSEGINVELRIGADLN 361
Db 312 FSLYAPHINSYKRFQPGSFAPTAVMGHDRNRTALRVV--GHGRSLRFENRLPGGDVNP 370
Qy 362 LAFSAITAAAGISGIEEKLPPASGNVYNDKELPEFPNSIQNATHLLKESKMLNKT 421

```

```

Db 371 LAVAGLVAAGLHGIEQRLPEPCPGNAYT-ADFAHVPTTLREAAELWENSTLAKAAGD 429
Qy 422 KLILHYVNAANVEINEFSKQVTDWELNQGFR 453
Db 430 EVVAHYRNMARVELDAFADAAVTDWELRRSFR 461

RESULT 8
Q936T0 PRELIMINARY; PRT; 459 AA.
AC Q936T0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Gamma-glutamylisopropylamide synthetase (EC 6.3.1.2) (Glutamine
DE synthetase) (Glutamate--ammonia ligase).
GN IPUC.
OS Pseudomonas sp. KIE171.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=159091;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KIE171;
RA de Azevedo Waesch S.I., van der ploeg J.R., Kiener A., Leisinger T.;
RT "Degradation of isopropylamine by Pseudomonas species KIE171.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + NH(3) = ADP + PHOSPHATE +
CC L-GLUTAMINE.
CC -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
DR EMBL; AJ311159; CAC81335.1; -.
DR HSSP; P06201; 2GLS.
DR InterPro; IPR001637; GlnA_adenyltn.
DR InterPro; IPR001691; Gln_synth.
DR Pfam; PF00120; gln-synt_1.
DR Pfam; PF03951; gln-synt_N; 1.
DR ProDom; PD001057; GlnA_adenyltn; 1.
DR Ligase.
KW Ligase.
SQ SEQUENCE 459 AA; 50586 MW; CBEP4392500A2436 CRC64;

Query Match 25.9%; Score 614.5; DB 2; Length 459;
Best Local Similarity 34.8%; Pred. No. 1.9e-42;
Matches 159; Conservative 73; Mismatches 194; Indels 31; Gaps 10;

Qy 7 ELNNLRNGKIDITVVLACVDMQGLMKRLTGRHFL---GLDQKISISFVAVTIEGI 63
Db 11 KVRDFIEKHNDITIRLGAVIDIGVWRGKQVGAEYFLNKAALDGTQISNILLFGWDVAH 70
Qy 64 AGGGEYISSVDTGYSDCHCADLNSLHLLPWSGAVLAISNPHNFTVTSPLFCSPRV 123
Db 71 --DGLFTGWDSGYPDIALIPDLSTLSLVPWQKTASVLCDIQI--LNGEPLNLSPRN 127
Qy 124 QOIERLANLKLGLFASLEFNLFNETYKSAQKHWNKLTAPQ---HHQWNIS 175
Db 128 KATKAEARQLGYKCAVAEVEFEYLLNDSIASISADQWRSINPVEKSGHCYMLHH----S 182
Qy 176 ASSGIETFMRSVRNKLREAGILMEATHPEFLPSQHELNFPADPLTMADRHIIAKH 235
Db 183 SSSDI---MGEVRKYRMDAGIVLEATNSEHGPQGYEINIKYDYDALKAAADAI 239
Qy 236 MAFSGMVATFMKLSLTALGNACHHMSIQDAETKNAFYQNDYQNDYQNDYQNDYQND 294
Db 240 IAAKHGTATFMKPSAEWSGSHVMSLSDLAGTP--VFAPENPCALSEVYNFLAG 297
Qy 295 LLKYVPEATYFFASVINSYKRLQPLTFAPTCCWADNRITSAFLCNKSKSEGINVEL 354
Db 298 MVALAREMSAIIPLNINSYKRTKAGASWAGGNSWGFNRTVSHRAITSAAGAA 357
Qy 355 GADLNPLYAPSAIIAGISGIEEKLPPASGNVYNDKELPEFPN----SLQNATHLLK 410
Db 358 GADTNPLYVIAAAILSLGSLGIENKLPKDPILGNAY--KVSPELARPLAASLEEA 415

```

QY 411 ESKMLNKTGFKLILHYVNAANVEINEFSKQVTDWEL 447  
 ||:| | : : || | : : || : : || : ||  
 Db 416 ESEKARVIFPNEFVEHYAQMKWEIKQSNFSFVNWEL 452

## RESULT 9

ID Q8GPV7 PRELIMINARY; PRT; 460 AA.  
 AC Q8GPV7;  
 DT 01-MAR-2003 (TReMBLrel. 23, Created)  
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
 DE Putative glutamine-synthetase.  
 GN GUNA4.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 NC NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SG17M;  
 RX MEDLINE=22313472; PubMed=12426355;  
 RA Larbig K.D., Christmann A., Johann A., Klockgether J., Hartsch T.,  
 Merk R., Wiehmann L., Fritze H.J., Tummeler B.;  
 RT "Gene Islands Integrated into tRNA(Gly) Genes Confer Genome Diversity  
 on a Pseudomonas aeruginosa Clone.";  
 RL J. Bacteriol. 184:6668-6680 (2002).  
 DR EMBL; AF40524; AAN62237.1; -. 35FD0D472764D446 CRC64;  
 SQ SEQUENCE 460 AA; 50620 MW; 35FD0D472764D446 CRC64;

Query Match 25.0%; Score 595; DB 2; Length 460;  
 Best Local Similarity 33.6%; Pred. No. 7.8e-41;  
 Matches 154; Conservative 80; Mismatches 191; Indels 34; Gaps 9;

QY 7 ELNNLRNGKIDTVLACVDMQGRMLGKRLTGRHFL---GLDQKKISITFFVAVT---I 60  
 ||:| | : : || | : : || : : || : ||  
 Db 11 KLOEFIEKHNDIVRLGAVDIDGVRGKQGAIFYLKAALDGTQISNIFLFGWDVADHLV 70  
 ||:| | : : || | : : || : : || : ||  
 QY 61 EGTAGGYEISSVDYTGSDCHLADLNSLHLLPWSGAVLAISNPNFVTSPLFCSPRV 120  
 ||:| | : : || | : : || : : || : ||  
 Db 71 DGLAFTGW-----DSGYFDIALIPDLTSLVLPWQAKTASVLCIDIQH-LNGQSLNLSPRN 124  
 ||:| | : : || | : : || : : || : ||  
 QY 121 ILMQOIERLANLKLGLFASLEFNFNFETYSKASQKHWKLNKTAOP-----HHQWM 172  
 ||:| | : : || | : : || : : || : ||  
 Db 125 LLKATKAKAQOLGYKCHAAFEFFYLLNDSIGSIADQWRSINPVKSGHCYSMLHH--- 181  
 ||:| | : : || | : : || : : || : ||  
 QY 173 NISAGSGITFMRSVRNKLSEAGILMEATHPEFLPSQHELNFPVADPLTMADRHIIAKHG 232  
 ||:| | : : || | : : || : : || : ||  
 Db 182 --SSSDI---IGEVRYMRDAGIVLEATNSEHGPGQYEINIKYDDALKAAADDAIFVKG 236  
 ||:| | : : || | : : || : : || : ||  
 QY 233 VREMAQSGMVATFMALKSLSTALGNACHIHMSLQDAETEKNAFYDQNDVEYGMSTLARNMI 292  
 ||:| | : : || | : : || : : || : ||  
 Db 237 IKEIAKHGATATFMAPNAWSGSGHVHISLNIITGAPAFANPENPAULSEVGYNFL 296  
 ||:| | : : || | : : || : : || : ||  
 QY 293 AGLLKTVPEATYFFASYINRYKRLQPLTFAPTRCCWAIDNRTSAFLCLNSKSGINVELR 352  
 ||:| | : : || | : : || : : || : ||  
 Db 297 AGNVELAREFSAYILPNINSYKKTAGASWAGNSWGFDNRTVSHRAITSAGAAARVENR 356  
 ||:| | : : || | : : || : : || : ||  
 QY 353 IGGADNPYLAFSAITAAAGISGIEEKLLEPPASGNVNDKELPEFPN-----SLQWATHL 408  
 ||:| | : : || | : : || : : || : ||  
 Db 357 IPGADNPYLVAASLLSGLYGIENLKKPKDPILGNAV--KVSPELARPLAASLEBATNI 414  
 ||:| | : : || | : : || : : || : ||  
 QY 409 LKESKMLNKTGFKLILHYVNAANVEINEFSKQVTDWEL 447  
 ||:| | : : || | : : || : : || : ||  
 Db 415 FRESEMARVLPFKPEFVEHYSQMKVWEIKQAGFVNWEL 453

## RESULT 10

Q9K665 PRELIMINARY; PRT; 455 AA.  
 ID Q9K665;  
 AC Q9K665;  
 DT 01-OCT-2000 (TReMBLrel. 15, Created)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)

DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)  
 DE Glutamine synthetase.  
 GN BH3867.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 NC NCBI\_TaxID=86665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RX MEDLINE=20512582; PubMed=11058132;  
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
 Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
 Horikoshi K.;  
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
 halodurans and genomic sequence comparison with Bacillus subtilis.";  
 RL Nucleic Acids Res. 28:4317-4331 (2000).  
 CC -1- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + NH(3) = ADP + GLUTAMINE +  
 ORTHOPHOSPHATE.  
 CC -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.  
 DR EMBL; AP001520; BAB07586.1; -.  
 DR HSSP; P06201; ILGR.  
 DR InterPro; IPR004809; GlnA.  
 DR InterPro; IPR001637; GlnA\_adenyln.  
 DR InterPro; IPR001691; Gln\_synth.  
 DR Pfam; PF00120; gln-synt; 1.  
 DR Pfam; PF03951; gln-synt\_N; 1.  
 DR ProDom; PD001057; GlnA\_adenyln; 1.  
 DR TIGRPFam; TIGR00653; GlnA; 1.  
 DR PROSITE; PS00180; GlnA\_1; 1.  
 DR PROSITE; PS00181; GlnA\_ATP; 1.  
 KW Ligase; Complete proteome.  
 SQ SEQUENCE 455 AA; 50463 MW; BB669371A49325F0 CRC64;

Query Match 21.5%; Score 510.5; DB 16; Length 455;

Best Local Similarity 31.1%; Pred. No. 8.1e-34;  
 Matches 145; Conservative 77; Mismatches 187; Indels 57; Gaps 12;

QY 6 DELNNLRNGKIDTVLACVDMQGRMLGKRLTGRHFLGLDQKKISITFFVAVTIEGIAG 65  
 ||:| | : : || | : : || : : || : ||  
 Db 16 EIKETIKTKVVELLHLOFVDIEGIL--KHVT-2-----VTAEQIDSVVD 56  
 ||:| | : : || | : : || : : || : ||  
 QY 66 GG--YEISVDYTG-----SDCHLADLNSLHLLPW-----SEGAVL-AISNPNFVTS 111  
 ||:| | : : || | : : || : : || : ||  
 Db 57 GKIMFDGSSI-TGFTPIQNSDLYLPDLTTFVAVLPWTEEGYSEGRFLCSVKKP---DG 111  
 ||:| | : : || | : : || : : || : ||  
 QY 112 EPLFCSPRVILMQOIERLANLKLGLFAS---ELEFNLFNETYSKASQKHWKLNKTAOPH 168  
 ||:| | : : || | : : || : : || : ||  
 Db 112 SDFEGDPRNLKKTVER---AKEKGYISVSGPELEFFLFE-----TDEYGLPTKOTDI 162  
 ||:| | : : || | : : || : : || : ||  
 QY 169 HQMMNISASSGJETFMRSVRNKLSEAGILMEATHPEFLPSQHELNFPVADPLTMADRHII 228  
 ||:| | : : || | : : || : : || : ||  
 Db 163 GGYFSPKDNCKEYKRLSEIYRTLKANGFTIEASHHEVAEGQHEINPKYDALGSDAAT 222  
 ||:| | : : || | : : || : : || : ||  
 QY 229 AKHGVRMAEQSGMVATFMALKSLSTALGNACHIHMSLQDAETEKNAFYDQNDVEYGMSTLA 288  
 ||:| | : : || | : : || : : || : ||  
 Db 223 YKWWVKTAKQPLHATFMPKPLGANGSGMHTNISLFDDEKQENAFYDESKLGLSKTA 282  
 ||:| | : : || | : : || : : || : ||  
 QY 289 RNIWAGLLKYPEATYFFASYINRYKRLQPLTFAPTRCCWAIDNRTSAFLCLNSKSGIN 348  
 ||:| | : : || | : : || : : || : ||  
 Db 283 YQFIAGLIENVKDFVAVTNPLVNSYKRLVPGVEAPCYTASNSRSLIRIPATRGAGR 342  
 ||:| | : : || | : : || : : || : ||  
 QY 349 VELRIGGADNPYLAFSAITAAAGISGIEEKLLEPPASGNVYN-----DKELPEFPNS 401  
 ||:| | : : || | : : || : : || : ||  
 Db 343 VEIRCPDPAANPYLAFVAVASAGLQGVQELTAPASVDADIFHMSAEETAEIGENLPTS 402  
 ||:| | : : || | : : || : : || : ||  
 QY 402 LQWATHLLKESKMLNKTGFKLILHYVNAANVEINEFSKQVTDWEL 447  
 ||:| | : : || | : : || : : || : ||  
 Db 403 LSEAINRFEGBIGKKTGFEHAFGEYVALKRGWEDEFRTTVHDWEV 448

## RESULT 11

Q9AEL4 PRELIMINARY; PRT; 427 AA.  
 ID Q9AEL4

AC Q9AEL4;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Glutamine synthetase 2 (EC 6.3.1.2) (Glutamate--ammonia  
 DE ligase).  
 GN GLNA2.  
 OS Corynebacterium glutamicum (Brevibacterium flavum).  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
 ON NCBI\_TaxID=1718;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC13032;  
 RA MEDLINE=21338478; PubMed=1145173;  
 RX Nolden L., Farwick M., Kraemer R., Burkovski A.;  
 RA "Glutamine synthetases of Corynebacterium glutamicum: transcriptional  
 RT control and regulation of activity";  
 RL FEMS Microbiol. Lett. 201:91-98(2001).  
 CC -1- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + NH(3) = ADP + PHOSPHATE +  
 CC L-GLUTAMINE.  
 CC -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.  
 DR EMBL; AJ310086; CAC34378.1; --  
 DR HSSP; P06201; ILGR.  
 DR InterPro; IPR001637; GlnA\_adenyltn.  
 DR InterPro; IPR001691; Gln\_synth.  
 DR Pfam; PF00120; gln-synt; 1.  
 DR Pfam; PF03951; gln-synt N; 1.  
 DR ProDom; PD001057; GlnA\_adenyltn; 1.  
 KW Ligase.  
 SQ SEQUENCE 427 AA; 48146 MW; C85326A81A2406F6 CRC64;

Query Match 19.8%; Score 470.5; DB 2; Length 427;  
 Best Local Similarity 28.9%; Pred. No. 1.5e-30;  
 Matches 134; Conservative 84; Mismatches 174; Indels 71; Gaps 12;

Qy 15 GKIDTVVLACVDMQGRMLGKRLTGRHFLGLDQKKISISTFYAVTIEGIAGGVEISVD 74  
 Db 10 GHLKSVVVAPELESAL-----EEGIGFD-----GSAIEGYA---RISEAD 47  
 Qy 75 TGYSDCHLCADLNSLHLLPWSG-AVLAINPHNFTV---SEPLFCSPRVILMQOIERLA 130  
 Db 48 T-----IARPDPTFQVLPLEAGISKLQAARLFCVDVMPDQGFSDPRQVLRQVQLAA 102  
 Qy 131 NLKLGIFASLEFLNFENETYSQSKHKNLKT-----AQPHHQWNISASSGIETFM 184  
 Db 103 DEGLTCHMSPIEFYLV-----QSLRTNGLPPVPTDNGGYFDQATFNEAPNFR 150  
 Qy 185 RSVRNKLEEEAGILMEATHPELPSQHELNFPADPLTMADRHIIAKHGVRERMAQSGMVA 244  
 Db 151 RNAMVALEELGIPVEFSHHTAPGQOEIDLRHADALTMAADNIMTFRYIMKQVARDQGVGA 210  
 Qy 245 TFMAKLSTALGNACHIMSLQDAETKNAPYDQNDYGVGMSLTARNWIAGLLKVVPSATY 304  
 Db 211 SFMPKPPQEHAGSAMTHMSLFEGDT--NAPHDPDDSYMLSKTAKQFIAGILHHAPEFTA 268  
 Qy 305 FFASYINSYKRLQPLTAPTKCCWAIDNRTSAFRLCN---SKSEGINVELRIGGADLNPNY 361  
 Db 269 VTQWVNSYKRIIVYVGNAPTAATMGVSNRSALVRVPTVRLNKESSRRVEVRLPDTACNPNY 328  
 Qy 362 LAFSAIIAAGISGIEEKLPPPPASGNV-----YNDKELPEFPNSLQNAATHLL 409  
 Db 329 LAFSVMLGAGLKGIEGYELDEPAEDDISNLSFRERRAMGYND-----LPSSLDQALRQM 383  
 Qy 410 KESKMLNKTGFKELILHYVNAANVEINEFSKQVTDWELNQGPN 452  
 Db 384 EKSELVADILGEHVFEEFLRNKREWRDYOQITPWEELRNLD 426

RESULT 12

Q8NNH4

ID Q8NNH4

AC Q8NNH4;

PRELIMINARY; PRT; 446 AA.

DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Glutamine synthetase (EC 6.3.1.2).  
 GN CGL2229.  
 OS Corynebacterium glutamicum (Brevibacterium flavum).  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
 ON NCBI\_TaxID=1718;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;  
 RA Nakagawa S.;  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP005281; BAB99622.1; --  
 DR InterPro; IPR001637; GlnA\_adenyltn.  
 DR InterPro; IPR001691; Gln\_synth.  
 DR Pfam; PF00120; gln-synt; 1.  
 DR Pfam; PF03951; gln-synt N; 1.  
 DR ProDom; PD001057; GlnA\_adenyltn; 1.  
 KW Ligase; Complete proteome.  
 SQ SEQUENCE 446 AA; 50381 MW; 30E54AC65E3B2CF1 CRC64;  
 Query Match 19.8%; Score 470.5; DB 16; Length 446;  
 Best Local Similarity 28.9%; Pred. No. 1.6e-30;  
 Matches 134; Conservative 84; Mismatches 174; Indels 71; Gaps 12;  
 Qy 15 GKIDTVVLACVDMQGRMLGKRLTGRHFLGLDQKKISISTFYAVTIEGIAGGVEISVD 74  
 Db 29 GHLKSVVVAPELESAL-----EEGIGFD-----GSAIEGYA---RISEAD 66  
 Qy 75 TGYSDCHLCADLNSLHLLPWSG-AVLAINPHNFTV---SEPLFCSPRVILMQOIERLA 130  
 Db 67 T-----IARPDPTFQVLPLEAGISKLQAARLFCVDVMPDQGFSDPRQVLRQVQLAA 121  
 Qy 131 NLKLGIFASLEFLNFENETYSQSKHKNLKT-----AQPHHQWNISASSGIETFM 184  
 Db 122 DEGLTCHMSPIEFYLV-----QSLRTNGLPPVPTDNGGYFDQATFNEAPNFR 169  
 Qy 185 RSVRNKLEEEAGILMEATHPELPSQHELNFPADPLTMADRHIIAKHGVRERMAQSGMVA 244  
 Db 170 RNAMVALEELGIPVEFSHHTAPGQOEIDLRHADALTMAADNIMTFRYIMKQVARDQGVGA 229  
 Qy 245 TFMAKLSTALGNACHIMSLQDAETKNAPYDQNDYGVGMSLTARNWIAGLLKVVPSATY 304  
 Db 230 SFMPKPPQEHAGSAMTHMSLFEGDT--NAPHDPDDSYMLSKTAKQFIAGILHHAPEFTA 287  
 Qy 305 FFASYINSYKRLQPLTAPTKCCWAIDNRTSAFRLCN---SKSEGINVELRIGGADLNPNY 361  
 Db 288 VTQWVNSYKRIIVYVGNAPTAATMGVSNRSALVRVPTVRLNKESSRRVEVRLPDTACNPNY 347  
 Qy 362 LAFSAIIAAGISGIEEKLPPPPASGNV-----YNDKELPEFPNSLQNAATHLL 409  
 Db 348 LAFSVMLGAGLKGIEGYELDEPAEDDISNLSFRERRAMGYND-----LPSSLDQALRQM 402  
 Qy 410 KESKMLNKTGFKELILHYVNAANVEINEFSKQVTDWELNQGPN 452  
 Db 403 EKSELVADILGEHVFEEFLRNKREWRDYOQITPWEELRNLD 445  
 RESULT 13  
 Q8FNL7;  
 ID Q8FNL7;  
 AC Q8FNL7;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DE Glutamine synthetase II (EC 6.3.1.2).  
 GN GLNA2 OR C2127.  
 OS Corynebacterium efficiens.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.



Search completed: December 17, 2003, 22:44:19  
Job time : 45 secs

